BIOACTIVE COMPOUNDS FROM NEW ZEALAND MARINE ORGANISMS

A thesis submitted

under the requirements

for the Degree of

Doctor of Philosophy in Chemistry

at the

University of Canterbury

by

Shangxiao Li

University of Canterbury

Christchurch

New Zealand

1998

This thesis is dedicated to my father, Li Jinan,

who died on 30th of December, 1994.

CONTENTS

List of Figures

List of Tables

Abbreviations

CBA	Carboxylic Acid (solid phase extraction cartridge)
COSY	Correlated Spectroscopy
DEPT	Distortionless Enhancement by Polarization Transfer
Dns-Am	Dansylated Amino acid
FABMS	Fast Atom Bombardment Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
HFBA	Heptafluorobutyric Anhydride
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
HRMS	High Resolution Mass Spectrometry
HSCCC	High Speed Countercurrent Chromatography
HSMQC	Heteronuclear Single and Multiple Quantum Correlation
HSQC	Heteronuclear Single Quantum Correlation
LC/MS	Liquid Chromatography/Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
NOBA	3-Nitrobenzyl Alcohol
PDA	Photo-Diode Array
PEI	Polyethylene Imine (solid phase extraction cartridge)
PFPA	Pentafluoropropionic Anhydride
ROESY	Rotating Frame Overhauser Enhancement Spectroscopy
RPHPLC	Reverse Phase High-Performance Liquid Chromatography
RRIMP	Roche Research Institute of Marine Pharmacology
TFA	Trifluoroacetic Acid
TFAA	Trifluoroacetic Anhydride
TOCSY	Total Correlated Spectroscopy

Abstract

Primary screening of recent collections from the sea around New Zealand revealed a wide range of bioactivities, including antitumour, antiviral and anti-HIV activities. Seven marine organisms, including sponges and an ascidian, were examined to identify the active compounds responsible for these three type of biological activities.

A chemical screening protocol was established for the dereplication and prioritization of antiviral natural product extracts, which was also applied to extracts with other bioactivities. A recent modification included HPLC analysis of the eluates from solid phase extraction cartridges, coupling the UV profiles of the corresponding peaks in the HPLC traces with MarinLit database. Extracts from two marine sponges which contained known cytotoxic compounds, pateamine and calyculins respectively, were used to test the efficiency of the modified protocol. These two types of active compounds were quickly identified by searching the UV data field in MarinLit.

HIV-inhibitory activity was recognised in both aqueous and organic extracts from a sponge *Chondropsis kirkii.* The aqueous extract was processed on a Sephadex-25 column. The positive results of a toluidine blue test on the resultant fractions suggested that sulphated polysaccharides may be responsible for the potent anti-HIV activity. The HIV-inhibitory components in the organic extract turned out to be a mixture of sterols. However, an increase of the cytotoxicity with increasing sample purity was also observed.

Examination of a sponge *Callyspongia irregularis* resulted in the isolation of a known compound, mycalamide A, which is a potent antiviral agent.

Purification of the trace amount of compound obtained was mainly achieved by high speed countercurrent chromatography (HSCCC). Examination of another *Callyspongia* species has led to the discovery of a potently cytotoxic fraction (IC_{50} 11 ng/mL).

A New Zealand deep water sponge, *Lamellomorpha strongylata*, has turned out to be a rich source of bioactive compounds. Bioassay-guided analysis led to the separation of five new cytotoxic theonellapeptolides IIIa, IIIb, IIIc, IIId and IIIe (molecular weight 1376-1453). These tridecapeptides were characterised by amino acid analysis using GC-MS, sequencing the ring-opened peptides with FABMS-MS, and 1D (^{1}H, ^{13}C, DEPT and decoupling experiment) and 2D (COSY, TOCSY, HMQC-TOCSY, HMBC, HSMQC and ROESY) NMR, IR analysis and chemical reaction. Of these, the MS-MS technique played a very important role in sequencing. These compounds exhibited weak cytotoxicity. However, the other member of the theonellapeptolide family showed ion-transport properties.

The absolute stereochemistry of IIIe was determined by X-ray crystal structure analysis in conjunction with a chiral HPLC method. The solution conformations of IIIb and IIIe were established mainly by correlations in the ROESY spectra. Six optically pure N-methyl amino acids were synthesised for a study of the stereochemistry of the component amino acids. The absolute stereochemistry of IIIa, IIIb, IIIc and IIId were determined by chiral LC-MS analysis. In addition, three minor components were isolated from the same sponge. ^{1}H NMR spectra showed that they all contained peptide components.

Acknowledgements

Above all I wish to thank my supervisors, Dr. John W. Blunt and Dr. Murray H. G. Munro, for their guidance, enthusiasm and encouragement, and for help in my life in Christchurch.

I am very grateful to Dr. Eric J. Dumdei for extraction of the sponge (*Lamellomorpha strongylata*) and for passing on the peptide fraction to me, as well as for his suggestions on structural elucidation. I especially thank Dr. Lewis K. Pannell (National Institute of Health, USA) for his excellent work on HRFABMS, MS-MS and LC-MS, and for his advice on peptide sequencing.

The technical assistance from both University of Canterbury and other institutes is gratefully acknowledged. Special thanks must go to Gill Ellis for the antiviral and P388 assays, to Bruce Clark for mass spectrometric and GC-MS analysis, to Professor Ward T. Robinson for obtaining crystallographic data, and to PharmaMar SA (USA) for the anti-HIV assay.

I also thank my lab-mates, particularly David J. Stirling, Sarah Hickford, Gill Nicholas, Rachel Lill, Michael Stewart and Joanne Hart, for their friendship and support.

I would also like to thank my wife for her patience and encouragement.

Chapter 1

Introduction

1.1 Marine Natural Product Chemistry

One of the basic principles of ancient Chinese philosophy was "integrity of nature and human body". It implied that the human body is a part of nature, or that the life and function of the body are closely associated with nature. In the long history of the human race, nature has not only offered the food and clothes for man's survival, but has also provided medicine for man's health. Modern science has demonstrated that the chemical constituents from plants, animals and microbes (natural products) are responsible for the medicinal benefits observed. Scientists have been successful in the development of drugs from natural products. In the last decade, over 60% of approved drugs and pre-NDA (pre-New Drug Application) candidates (for the period 1989-1995) are of natural origin.[1] Natural products from marine organisms may have considerably greater potential than those from terrestrial habitats in terms of pharmacological uses.[2] The oceans cover about 71% of the earth's surface, and have a wide biodiversity. An estimated range of total global biodiversity is between 3 million and 500 million different species. Approximately a half of these species live in the oceans.[3] All but 2 of 28 major animal phyla are represented in aquatic environments, with 8 being exclusively aquatic, mainly marine.[4] An equally important factor in terms of diversity of secondary metabolites is the unique biosynthetic conditions of the marine environment. Marine organisms live in a different ionic strength environment, with higher pressure and lower temperature conditions than their terrestrial counterparts. This combination of factors may produce some compounds which would be unstable under other conditions.[5] The marine world is a very competitive and hostile environment; thus many marine organisms have developed complex attack

and defence mechanisms, as well as symbiotic associations, to enhance their survival capability.[6] However, despite successes in recognising bioactivity from marine resources and the belief that the marine environment is an exceptional reservoir of bioactive natural products, the research into the use of marine natural products as pharmacological agents can still be considered to be in its infancy.[7]

In its early phase, the research on marine natural products was driven by chemical studies,[7,8] which were serendipitous studies,[9] although Bergmann's pioneering work in 1950 resulted in the development of two antiviral and antitumour drugs. A realisation of the biomedical potential of marine natural products began with the first in a series of Symposia on Food and Drugs from the Sea in 1969.[10] The first marine biomedical workshop was held in 1974 at the Smithsonian Institution in Washington. A variety of problems in marine biomedical research and their possible solutions were reviewed. This resulted in a paper identifying the major, future milestones and directions of research in the pharmacological-biomedical exploration of the sea.[11] The first systematic investigations of biologically active marine organisms began in the 1970's.[12] The R/V ALPHA HELIX Baja expedition in 1974 assayed 850 marine species for antimicobial and antifungal activities, and in 1978 an antiviral investigation of Caribbean marine organisms was carried out, which culminated in the first clinical trial of a marine natural product (didemnin B).[13] In 1974, the Roche Research Institute of Marine Pharmacology (RRIMP) was established for the commercial development of pharmaceuticals from marine resources.[14] In its seven year operational period, RRIMP identified more than 400 novel compounds from sponges, corals, echinoderms, molluscs, ascidians, sea grasses, microalgae, bacteria, algae, fish and other marine organisms. The RRIMP screens

included *in vivo* evaluation in gross observation, toxicity, effects on central nervous system, anti-inflammatory, anti-allergy, analgesia, cardiovascular effects, neurophysiology and biochemical pharmacology.[14]

Although the commercial exploitation of new drugs from the sea was up and down,[15] interest in the development of marine natural products increased rapidly. Between the 1940's and 1950's, only a few dozen papers were published. The number of publications was in the hundreds in the 1960's, and jumped to the thousands in the 1970's. This increasing rate was maintained during 1980's, with more than 3,000 more papers published.[15] Meanwhile, the number of new compounds reported increased at the rate of 10% per year.[16] So far, more than 7,000 compounds have been described.[17] The studies have been extended to all phyla of the marine organisms. However, the distribution of the compounds in the phyla is not even. The majority of the compounds have been isolated from sponges (33%), algae (25%) and coelenterates (18%), with the balance 24% coming from other invertebrate phyla, such as tunicates, molluscs, echinoderms and bryozoans.[16]

Since the 1970's, research in marine natural products has been focused on pharmacologically and biomedically significant compounds. A wide range of screening methods have been utilised. Looking at the literature, a profound portion of the secondary marine metabolites in the earlier studies showed antibacterial activity, while antitumour compounds have been dominant more recently[18]. During the 1960's and 1970's, the antibacterial assay method was well established and most readily available to chemists. This can explain the high incidence of the observation of antibiotic activity. In the last two decades, the cytotoxicity bioassay method was widely adopted by chemistry laboratories due to its simplicity

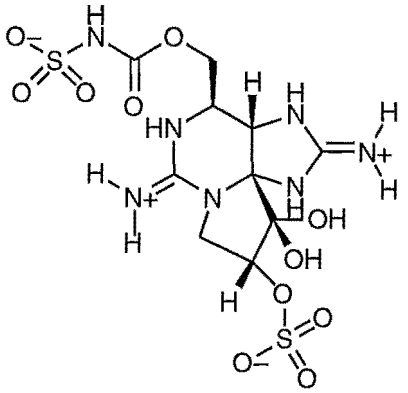
and relatively low cost. At the same time, both government and industry have put more and more funding into cancer research. All these factors have contributed to the discovery of the large number of cytotoxic compounds. The high incidence of cytotoxicity of marine natural products may also be considered an outcome of the evolutionary history of the marine organisms.[16] Some of them, for example sponges, are at a very primitive stage. Other frequently used biological tests include screening for cardiovascular, antifungal, antiviral, anti-inflammatory, immunosuppressive, antihelminthic, central nervous system, hypothermic and analgesic activity. In recent years, increasing interest in mechanism-based screening of natural products has been observed due to rapid development of biotechnology and a greater industrial involvement.

1.2 The Chemical Ecology of Marine Organisms

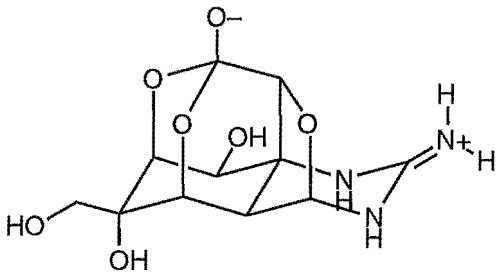
In the course of developing biomedical compounds from marine sources, there are two important questions that have attracted the attention of both chemists and biologists for a long time. i) What is the true origin of the secondary marine metabolites? ii) What are the biological roles of these secondary metabolites in the marine organisms?

To date, many suggestions have been made as to the sources of the biologically-active compounds.[20] Accordingly, the bioactive natural products may come from micro-organisms associated with the marine macro-organisms, including symbiosis and epibiosis,[21] or from the food chain by feeding. Recent studies have suggested that some bioactive compounds isolated from marine invertebrates, such as sponges, coelenterates, molluscs, or protochordates, are actually generated by microorganisms.[22,23] For example, saxitoxin (**1**), a well-know shellfish poison, comes from dinoflagellates of the genus *Gonyaulax* and the blue-green alga *Aphanizomenon flos-aquae.*[24] Tetrodotoxin (**2**), originally isolated from the pufferfish,[25] has also been found in a red alga[26] and a xanthid crab.[27] More typical examples are found with sponges. Structurally identical or similar metabolites have been isolated from different species of sponge; while in other cases, the same species contained totally different compounds.[28] The association of microbes with sponges has been well documented.[21] The microorganisms found in sponges are predominantly bacteria[29] and cyanobacteria (blue-green algae), and to a less extent dinophycean *Zooxanthellae* and eucaryotic algae.[30] More direct evidence for the symbiotic origin of sponge metabolites has been obtained for the sponge *Tedania ignis*, in which

several diketopiperazines were identified.[31] When several species of the genus *Micrococcus* bacteria were isolated from the sponge and cultured in the laboratory, the bacteria were found to produce the same diketopiperazines as those obtained from the sponge. This suggested that these compounds were most likely bacterial metabolites.[32]

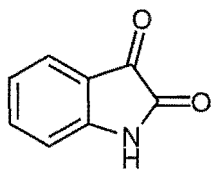


Saxitoxin (**1**)

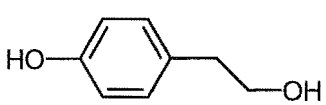


Tetrodotoxin (**2**)

When the microorganisms live on the surface of macro-organisms, such as sponges, corals and macroalgae, they are called epibionts. Epibionts fulfil important roles (eg. antifouling, antifungal, antibacterial and settlement aiding activities) for the host organisms. The embryos of the crustaceans *Palaemon macrodactylus* and *Homarus americanus* are covered with a unibacterial film. On fermentation, these bacteria yielded the antifungal compounds 2,3-indolinone-dione (**3**) and tyrosol (**4**), and it was demonstrated that these compounds prevented fungal infection of the embryos.[33, 34]



2,3-Indolinone-dione (**3**)



Tyrosol (**4**)

Faulkner[20] has pointed out that unless a symbiont is observed, food-chain accumulation and convergent evolution must be considered as alternatives to explain the origin of compounds previously assigned to "symbiosis".[20] A good example of food-chain origin of the biological metabolites is found with the nudibranchs, which are specialist sponge grazers. While feeding on sponges, the nudibranchs sequester the effective chemical armoury of their prey, which is subsequently employed for their own protection.[34]

The biological roles of secondary metabolites are normally related to antipredation, antifouling, competition for space, or species dominance etc.[35] Marine natural products that have been related to chemical ecology include terpenes, steroids, phenolics and nitrogenous compounds.[35] The high incidence of toxic, or deterrent sponge metabolites is found in habitats such as coral reefs, where the competition for space and feeding pressures are very high. Another piece of evidence for the adaptive significance of sponge constituents is the observation that sponges growing in exposed areas are usually more toxic than those found growing in less exposed situations. A striking example of such chemical defence is provided by a Red Sea sponge, *Latrunculia magnifica*. When the juice of this sponge was squeezed into an aquarium stocked with the fish *Gambas afghanis*, all the fish died within a few minutes.[36] The occurrence of clean-surfaced animals that lack appendages, such as sponges, ascidians, soft corals, gorgonians and holothurians, suggests that antifouling allomones may be present. Marine organisms, in competing for space, may crowd, undercut, crush, overgrow, digest or poison their neighbours.[37] Soft corals may move over hard corals and kill them to take space susceptible to secondary colonisation.[38] Do all these biological functions have chemical bases?

A proper answer to the above questions is important for the study of the biosynthesis of active metabolites. Furthermore, the information might even be vital for the economic production of pharmaceuticals from marine sources.[20] Moreover, it may provide guidance to allow natural product chemists to locate further marine organisms containing bioactive compounds.

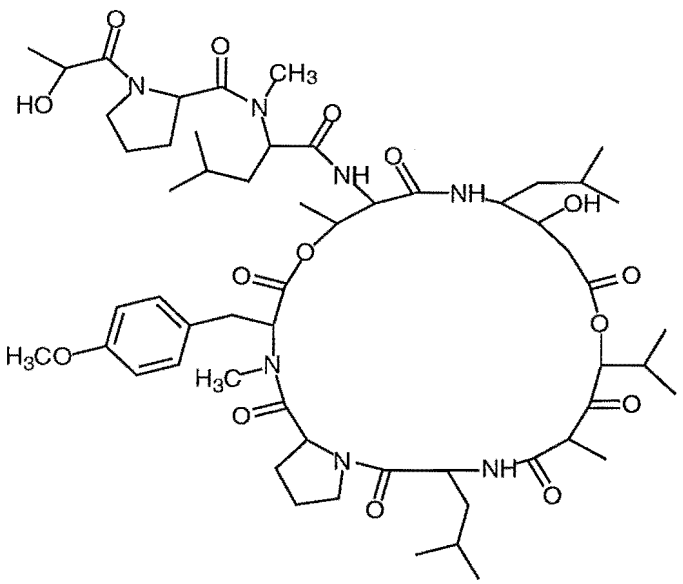
1.3 Commercial Uses and Potential Applications of Marine Natural Products

1.3.1 Drugs from the Sea

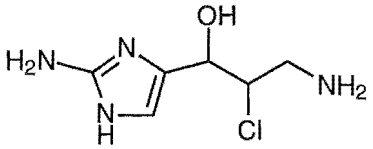
Living a healthier and longer life has been the dream of human beings. Historically, natural products from terrestrial plants and micro-organisms have played vital roles in the development of new drugs.[39] Thus, it is not surprising that marine natural product chemists focus their attention mainly on the biomedical applications of the secondary metabolites from marine sources. Although the potential of marine natural products as a new drug source was realised as early as 1969,[10] no new drugs derived directly from marine organisms have been marketed to date. However, the enthusiasm for biologically active compounds from the sea has not faded. With more in-depth studies, even more potential in this field can be recognised since a higher incidence of bioactive compounds has been recognised from marine organisms than those encountered from terrestrial sources.[40] (A marketable new drug is probably on the verge of breakthrough.)

Currently, there are several marine related compounds at a clinical, or preclinical testing stage. The first marine natural product which progressed to clinical trials was didemnin B (**5**). This depsipeptide was isolated from a Caribbean ascidian *Tridemnin solidum* by Rinehart *et al.*[41] This compound exhibited a wide range of biological activities, such as antiviral, immunosuppressive, and cytotoxic activity. It is clinically tested only as an anticancer drug due to its highly cytotoxic properties. Another early marine compound that proceeded towards clinical trials was girolline[15] (**6**), but testing was stopped due to blood-pressure

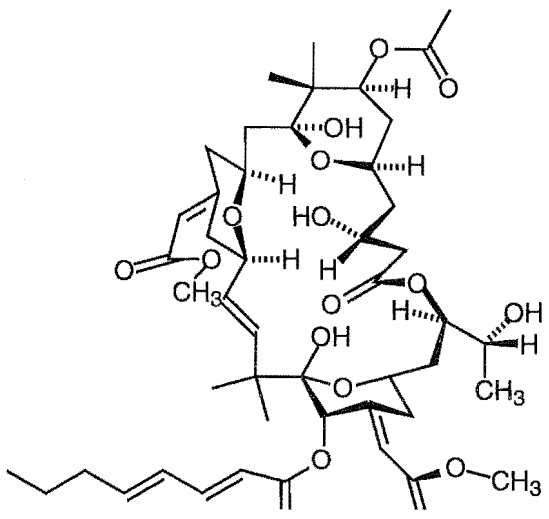
problems in the patients. Another anticancer compound in clinical evaluation is bryostatin 1 (**7**).[42] Dolastatin 10 (**8**), a peptide isolated from the sea hare *Dolabella auricularia*,[43] is in phase I clinical trials. Ecteinascidins (**9**), alkaloids isolated from the tunicate *Ecteinascidia turbinata*,[44] are nearing phase II clinical trials.[21] Another two cytotoxic compounds, halomon (**10**) and halichondrin B (**11**), are in preclinical evaluation. The former from the red alga *Portieria hornemannii* exhibits promising differential cytotoxicity against several tumour cell lines.[45]



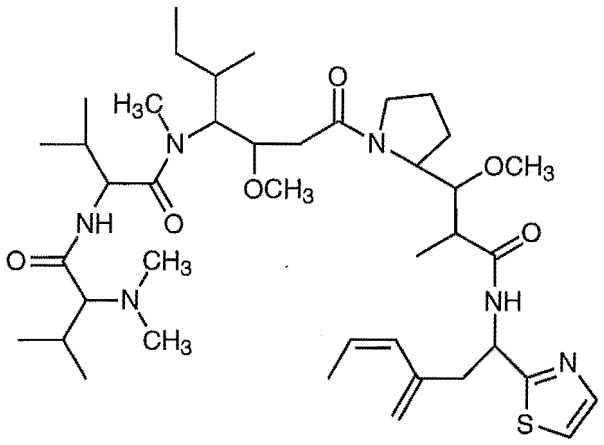
Didemnin B (**5**)



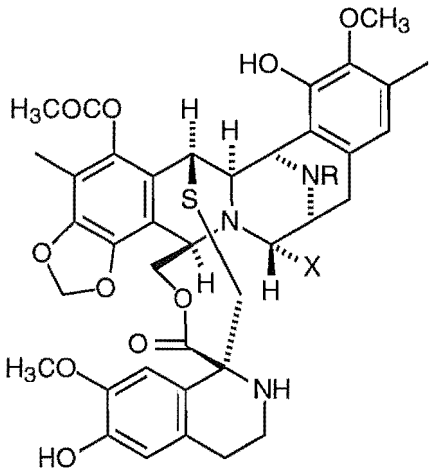
Girolline (**6**)



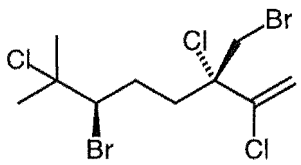
Bryostatin 1 (**7**)



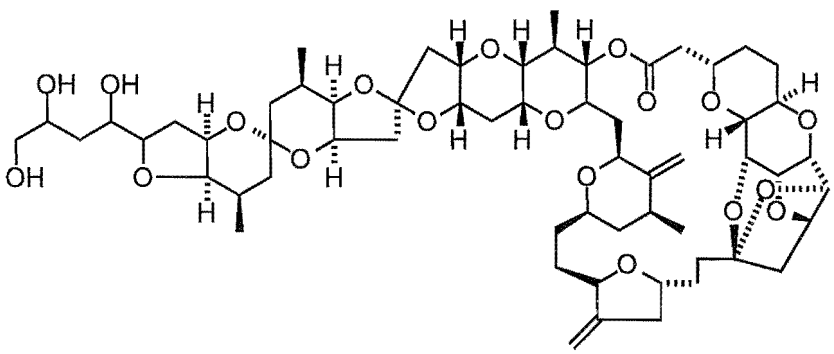
Dolastatin 10 (**8**)



Ecteinascidins (**9**) R= H or CH_3
X= OH, OCH_3, CN



Halomon (**10**)



Halichondrin B (**11**)

The latter was first isolated from a Okinawan sponge,[46] and later obtained from a New Zealand deep water sponge *Lissodendoryx* sp.[47]

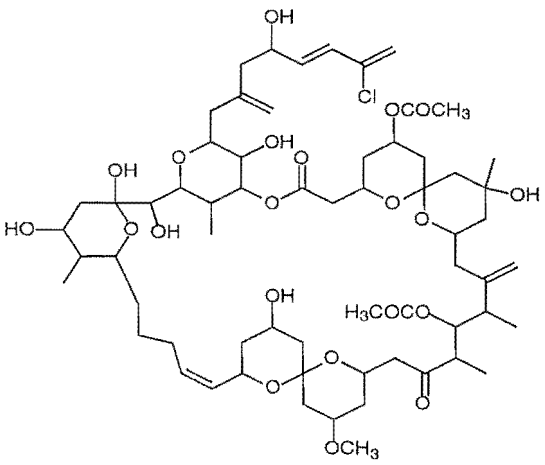
An Eastern Indian Ocean sponge in the genus *Spongia* produced a structurally unprecedented marocyclic lactone, spongistatin 1 (**12**), with extremely potent cytotoxicity against human tumour cell lines.[48] Spongistatin 1 exhibited GI_{50} values of 2.5-3.5 x 10^{-12} M against several highly chemoresistant tumour types, eg. HL-60, SR leukemias, NCI-H226, NCI H23, NCI H460, NCI H522 non-small cell lung, DMS 273 non-small cell lung, HCT-116, HT 29, KM 12, KM 20L2 SW-620, SF-539, U-251 brain, SK-MEL-5 melanoma, OVCAR-3 ovarian and RXF-393. Spongistatin 1 can effectively inhibit the cell lines derived from human melanoma and lung, colon and brain cancer.

There are two anti-inflammatory compounds, which have progressed to clinical trial stage. Manoalide (**13**), isolated from the sponge *Luffariella variabilis*,[49] exhibited inhibiting activity against phospholipase A2 (IC_{50}=0.05 μM) and phospholipase C (IC_{50}=1.5 μM), resulting in reduced arachidonic acid release. This compound was also applied to the study of the role of calcium-channel blocking agents and signal transduction pathways due to its calcium-channel blocking activity. A series of derivatives of manoalide have been synthesised, some of which have been clinically evaluated.[50]

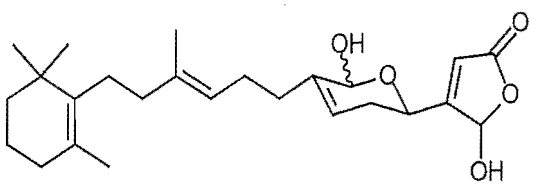
Anti-inflammatory depsipeptides, salinamides (**14**) A and B, were found from a marine *Streptomycete*. These two peptides have non-amino acid fragment containing an epoxide and chlorohydrin respectively. The most potent *in vitro* activity is against *Streptococcus pneumoniae* and *S. pyrogenes*. Importantly, they exhibited potent topical anti-inflammatory activity against the phorbol ester-induced mouse ear edema assay.[51]

Since ara-A and ara-C, the first antiviral agents, derived from marine ara-nucleosides were developed, discovery of antiviral drugs from the sea has been one of the main targets of marine natural product chemists. The results are very encouraging. A wide range of marine secondary metabolites showed antiviral activity. Examples include reiswigins (**15**) from *Epipolasis reswigi*, isospongiadiol from a *Spongia* sp., topsentins (**16**) from a Mediterranean *Topsentia genetris* and a Caribbean *Halichondria* sp., and sulfolipids, anti-HIV compounds from a cyanophyte.[15]

Modern medicinal studies have revealed that disorders of the immune system may cause a variety of diseases, such as diabetes, arthritis, psoriasis, multiple sclerosis, as well as the rejection problems in organ and tissue transplantations. Immunoregulating drugs have a great potential for clinical application. Marine natural products turn out to be a very promising source of immunoregulating agents, although only a very small portion of marine organisms have been tested for this kind of activity. Of them, discodermolide (**17**) and microcolin A (**18**) exhibited very promising immunosuppressive properties. Discodermolide, obtained

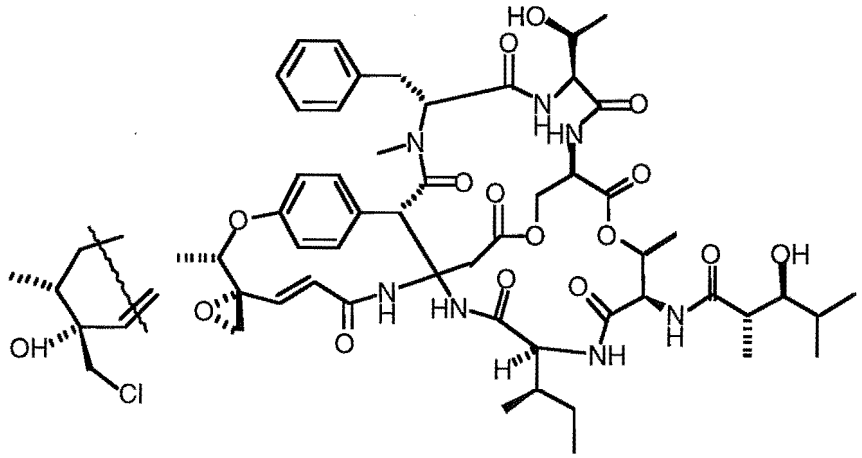


Spongistatin 1 (**12**)

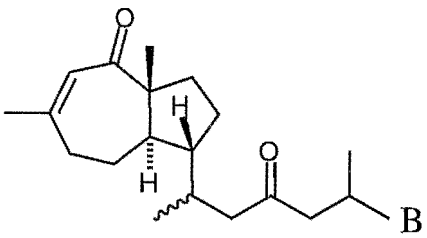


Manoalide (**13**)

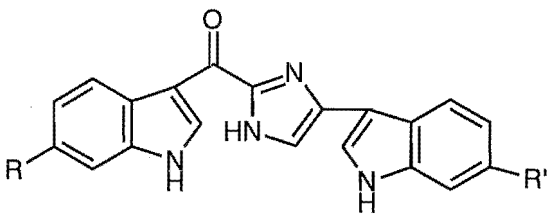
from *Discodermia* sp., is a polyhydroxylated lactone.[52] It has a different mechanism of action from the currently used immunoregulating drug, cyclosporine A. The former showed suppressive activity in the murine two-way MLR and conconavalin A stimulation of splenocyte cultures, and the proliferation of human peripheral blood leukocytes at relevant concentrations (<80.6 μM).



Salinamides (**14**)



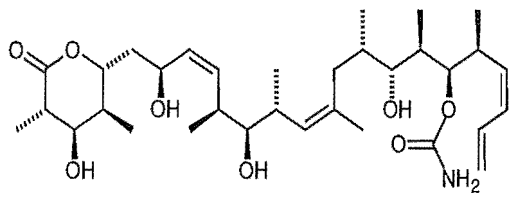
Reiswigins (**15**)



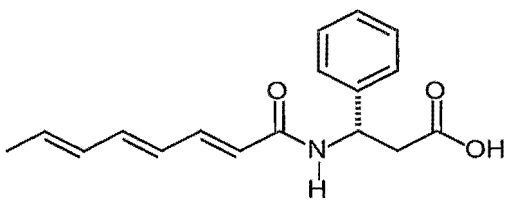
Topsentins (**16**) A, R=R'=H
B, R=OH, R'=H
C, R=OH, R'=Br

As well as the therapeutic targets mentioned, a series of other pharmacological models are also applied in screening bioactive marine natural products. A wide diversity of biological activities, such as antibacterial, antifungal, cardiovascular, antihelminthics, antithrombic and multidrug-resistance reversing activity, have been explored. In an attempt to find new drugs against antibiotic-resistant pathogens, the

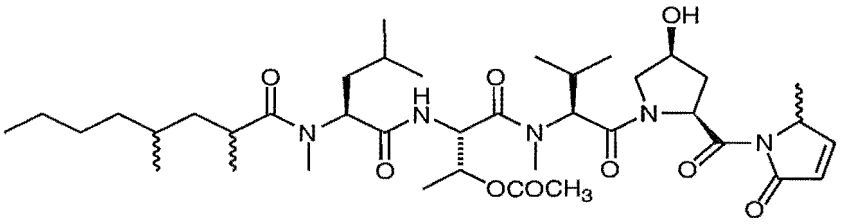
bacterium *Pseudomonas fluorescens* was isolated from a tunicate. After cultivation on solid media, two potent antibacterial compounds, andrimid (**19**) and moiramide B (**20**), were identified.[53] The former was previously obtained from an *Enterobacter* sp.. During chemotherapy, a frequently encountered problem is that cancer cells become resistant to anticancer drugs when a drug is used for long term treatment. A cyclic depsipeptide, hapalosin (**21**), isolated from a blue-green alga (cyanobacteria) *Hapalosiphon welwitschii*, exhibited multidrug-resistance reversing activity.[54] This compound can potentiate the cytotoxicity of common antitumour drugs like vinblastine and adriamycin toward drug-resistant cells. A Fijian Jaspidae sponge was found, containing anthelminthic components, the bengamides (**22**).[55] Cyclotheonamides (**23**), isolated from a sponge *Theonella* sp., are potent thrombin-inhibiting cyclic peptides.[56] Moore *et al.*[57] discovered a cyclic peptide, puwainaphycin C (**24**), from a blue-green alga *Anabaena* BQ-16-1. This peptide elicits a strong, positive inotropic effect in isolated mouse atria. Gambieric acids A-D (**25**) were obtained from the cultured marine dinoflagellate *Gambierdiscus toxicus*. These polyether metabolites are the most potent *in vitro* antifungal agents reported to date. Only moderate toxicity against mice or cultured mammalian cells was observed for these compounds.[58] They will possibly be developed into new antifungal drugs.



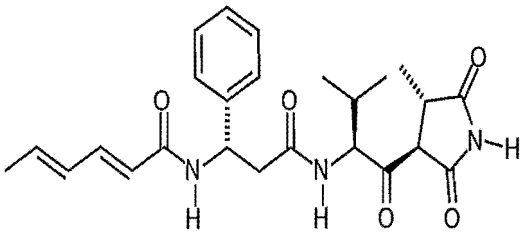
Discodermolide (**17**)



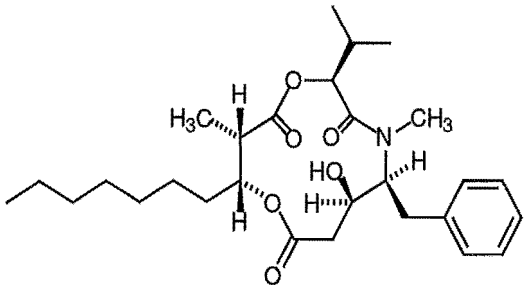
Andrimid (**19**)



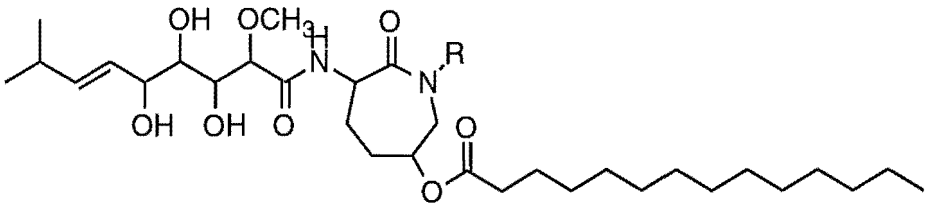
Microcolin A (**18**)



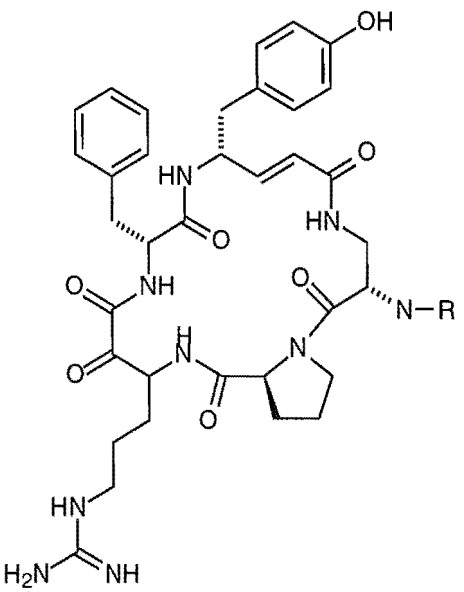
Moiramide B (**20**)



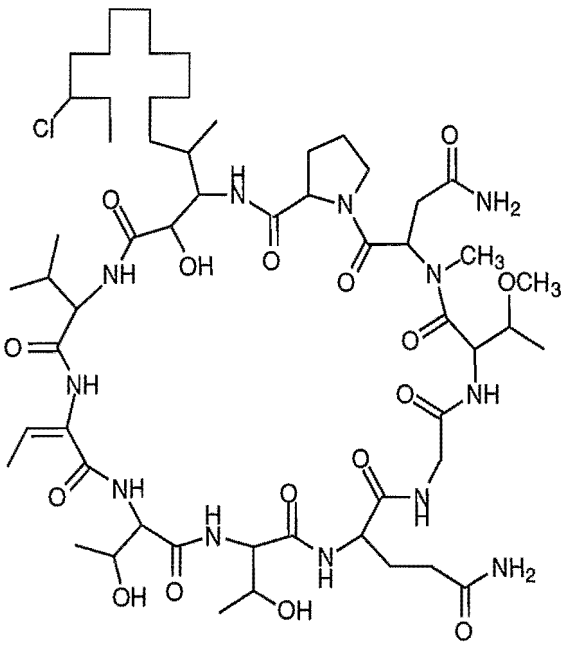
Hapalosin (**21**)



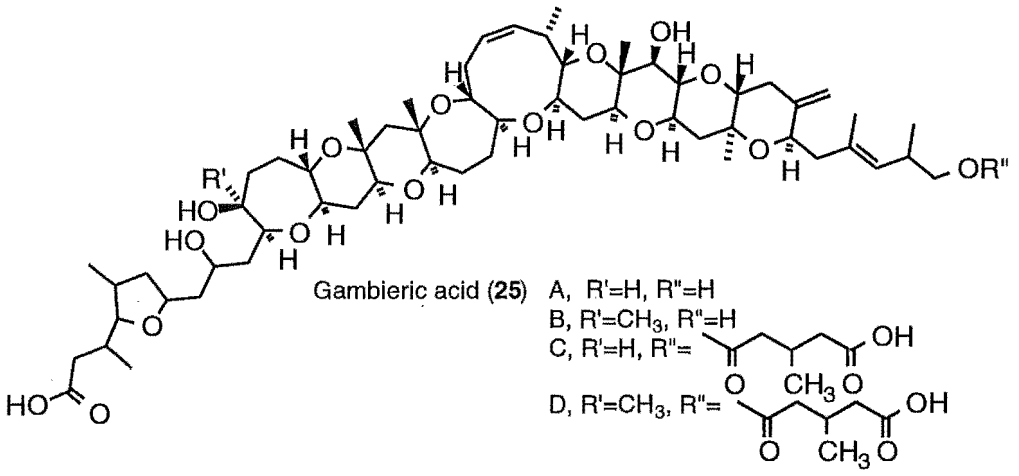
Bengamides (**22**) A, R=H, B, R=CH_3



Cyclotheonamides (**23**) A, R=CHO
B, R=Ac



Puwainaphycin C (**24**)



1.3.2 Biomedical Tools

Tetrodotoxin (**2**) was originally isolated from the liver and gonads of a certain puffer fish (*Tetraodon* group).[59] The toxin was also found in the porcupine fish (*Diodontidal*), the ocean sunfish (*Molidae*), newts (*Taricha*), atelopia, the blue ringed octopus, several species of biota and bacteria, as well as some South American frogs. It is about 160,000 times more potent in blocking axonal conduction than cocaine.[60] The main action of tetrodotoxin is paralysis of the peripheral nerves. It is widely used as an inhibitor in neurophysiological research since it inhibits specifically the sodium ion permeability of the membrane.[61]

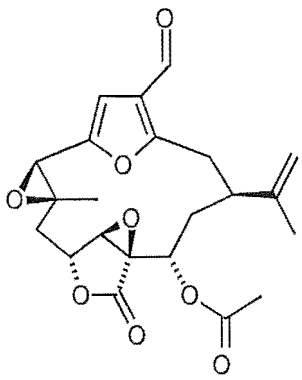
Lophotoxin (**26**), a paralytic pyran aldehyde derivative, was isolated from several species of the genus *Lophogorgia*. In receptor studies with lophotoxin, it was found to irreversibly inactivate the nicotinic acetylcholine receptor on intact BC3H-1 cells in culture. Since the binding is irreversible, it differs from all known ganglion blockers. This toxin has been used as an important neuropharmacological probe in studying complex neural pathways.[62]

Palytoxin (**27**) is an extremely poisonous, water soluble substance from marine cnidarians (*Palythoa*).[63,64] It influences calcium- and potassium-

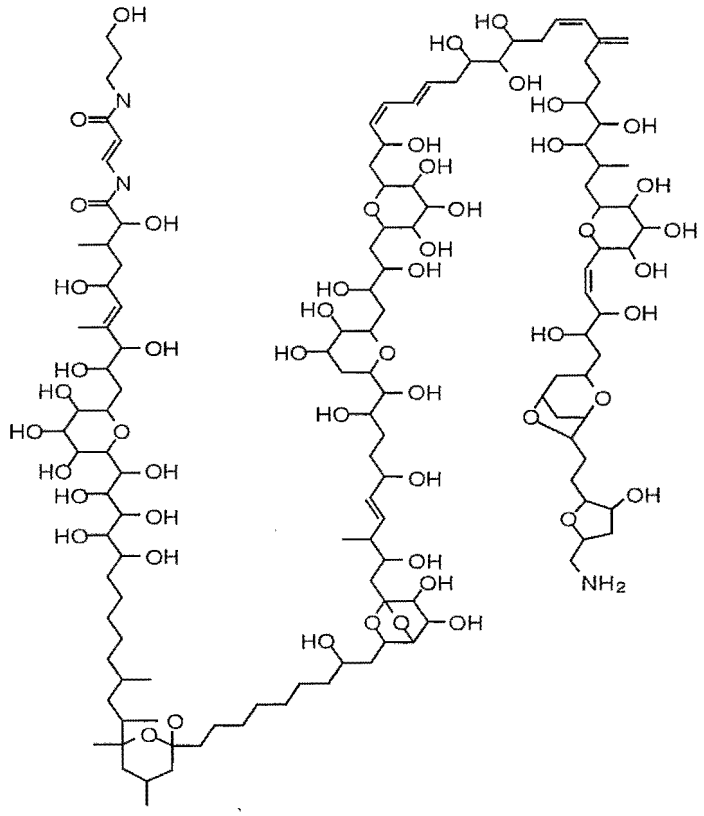
ion transport in nerve cells and heart tissue. Animals undergo paralysis and heart failure. Palytoxin has become a tool in ion transport studies.

Ciguatoxin (**28**) is a group of structurally similar, lipid-soluble compounds. It was originally recognised from human disease caused by the ingestion of any of a wide variety of coral reef fishes containing toxins. However, the true origin of the compound turned out to be a photosynthetic benthic dinoflagellate, *Gambierdiscus toxicus.* The toxins increase permeability of excitable membranes to sodium, causing depolarisation.

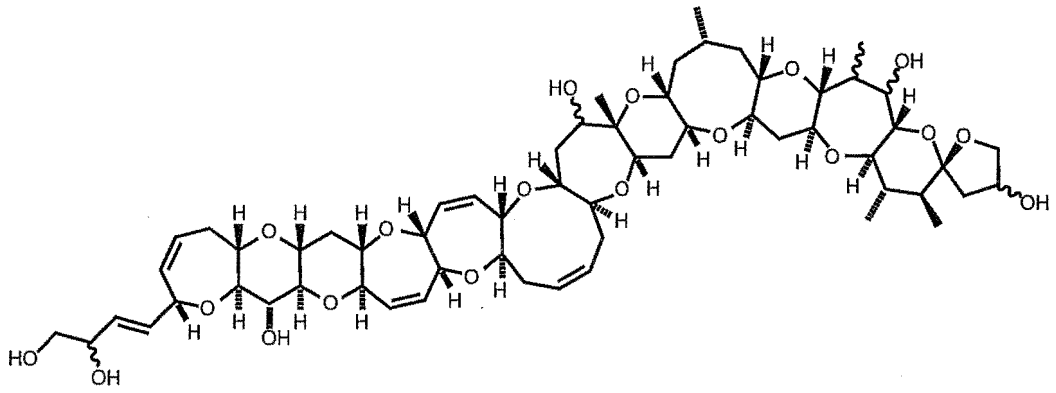
Okadaic acid (**29**) and related compounds are polyether toxins, which were first obtained from *Halichondria okadai* and *H. melanodocia.*[65] Later on they were also isolated from several dinoflagellates[66] and most



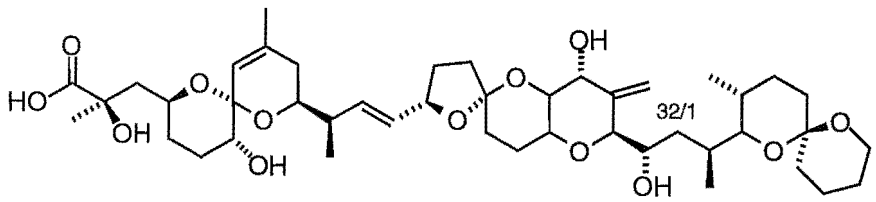
Lophotoxin (**26**)



Palytoxin (**27**)



Ciguatoxin (**28**)



Okadaic acid (**29**)

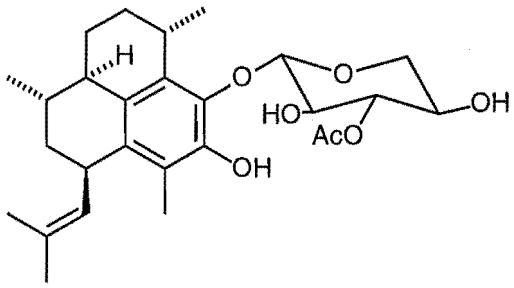
recently from a New Zealand sponge *Raspailia agminata.*[67] Okadaic acid and its congeners showed protein inhibiting, smooth muscle contraction affecting, and tumour promoting activities.[68]

1.3.3 Healthcare and Health-food

Trans/cis-β-carotene can be obtained from a variety of species of the genus *Dunaliella dunaliella* predominates in many aquatic marine habitats and in salt water bodies which contain more than 10% salt. *Dunaliella* is probably the most halotolerant eukaryotic organism known, showing a remarkable adaptation to a variety of salt concentrations from as low as 0.2% to about 35%.[69] It is believed that β-carotene possesses an antioxidant function and the ability to quench various radical species, which in turn may reduce the incidence of several types of cancer. As a food-colouring agent and as provitamin A, β-carotene is added to commercial products, such as cosmetics, health-foods, vitamin

preparations and pharmaceuticals. These products are marketed worldwide.[70]

A recently developed skin-care cosmetic contains the diterpene glycoside pseudopterosin C (**30**),[71] isolated from the gorgonian coral *Pseudopterogorgia elisabethae* collected in the Bahamas. This product is now marketed under the name Resilience by Estee Lauder in New York.[72]



Pseudopterosin C (**30**)

Eicosapentaenoate and docosahexaenoate, obtained from fish oil, exhibited activities in the prevention and cure of thrombosis and lowering of blood cholesterol.[73] Health food containing these compounds has been successfully marketed in United States, Japan and China.

1.3.4 Other Industrial Applications

Tyrian purple, the vibrantly coloured dye extracted from *Murex* sp. and related molluscs about 1600 BC, was probably the first commercial marine natural product.[74] The modern industrial applications of the materials derived from marine sources are mainly the polysaccharides. Agars, carrageenans and alginates are the best known and most widely used products of seaweeds. Agar is mainly prepared from several species of the genus *Gelidium*. It is frequently used in the pharmaceutical industry, such as the culture base for bacteria and a binding for tablets. It is also a laxative.[2] Carrageenan, another red seaweed polysaccharide

originally obtained from Irish Moss and *Chondrus crispus*, can be produced from several other genera now. A refined extractive is used for emulsions, eg. in a cough syrup and in tooth pastes. Its potassium salt is utilised in shaving soaps and haircreams, and as a binding agent for tablets.[75] It was also found to be valuable in the control of ulcers through action with the mucous lining of the stomach.[76] Sodium alginate, an excellent suspensory agent, has had extensive applications in both the pharmaceutical and food industries. Algin capsules are usually used to deliver the drugs that must be liberated intestinally, because the alginates are only hydrolysed in the intestine. There are a number of dental uses, eg. in moulds and the control of bleeding.

A successful commercial development of a marine natural product is the insecticide derived from nereistoxin (**31**). This toxin was first isolated from *Lumbrinereis brevicirra* in 1934. Once its structure was established, cartap hydrochloride, one of its derivatives, was synthesised. It is active against the rice stem borer and other insect pests, and does not appear to be toxic to warm-blooded animals.[77] This product has been widely used in agriculture since 1966.

1.4 Problems and Challenges in Marine Natural Product Studies

The collection of marine organisms is a difficult task. It requires both wide biological knowledge and special equipment. The usual approach is by SCUBA diving. It is hard for the divers to determine exactly where they should collect the specimens, because visibility is often poor under water. It is impossible to know the abundance and variety of animals in a particular habitat before the divers get to the bottom. As a result of the wind and waves, the collection sites are frequently obscured. Recollection of the biologically active organisms is even more difficult, because the record for original collection is less accurate, or the bioactive components vary with seasonal and geographical factors. Marine natural product chemists often have trouble in actually identifying the specimens due to the worldwide shortage of taxonomists.

Usually, the biologically active components exist in very low concentration in the marine organisms. It is often a challenge for the natural product chemist to obtain enough of the pure compounds for chemical and biological study. Some of the marine natural products are sensitive to light, moisture, or oxygen, especially as they become purer during the isolation process. Separation of these subtle compounds is frustrating work. In the extreme situation, several years of hard work may be required to obtain enough of the pure compound to establish the structure.[78]

The good old days of 'grind and find', when nearly every organism contained new and interesting molecules, are fast disappearing. The most bioactive extracts now mainly contain only known compounds.[16] Avoiding the duplication of work on known active compounds is

important both for saving funding and time, and for patent application. Development of rapid methods to gain insight into the properties of the active compounds before initiation of the purification procedures has become a priority.

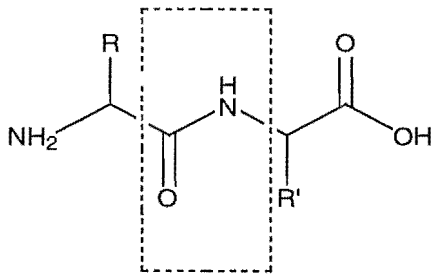
Another challenge is supplying enough pure material for pharmacological study and commercial development of the new drugs. Naturally growing marine organisms may either have limited biomass available, or contain very low concentrations of the active components. The structures of the biologically active compounds sometimes are too complex to be synthesised economically. This problem may hinder the further development of a new drug. Some solutions have been proposed for these problems.[15] These include collection in large quantities but without damage to the environment from the collection, organic synthesis, aquaculture, tissue culture and gene engineering techniques. Of these options, aquaculture shows the most promise. Two recent examples give very encouraging results. One is aquaculture of a colonising bryozoan, *Bugula neritina,* for the large scale production of the anticancer agent bryostatin 1. CalBioMarine Technologies has successfully cultured this animal offshore of La Jolla, CA, and partially purified bryostatins from both wild and aquacultured materials.[79] Another example is culturing *Lissodendoryx sp.* for the anticancer compound, halichondrin B. To obtain enough pure material for pharmacological and preclinical studies, both a large-scale recollection and aquaculture were considered. The University of Canterbury group and NIWA, with National Cancer Institute support, has successfully transported sponge samples from depths of 70-100 m to 10-30 m depth, while maintaining production of halichondrins.[2] These results demonstrated the feasibility of commercial production of bioactive compounds by aquaculture.

1.5 Marine Peptides

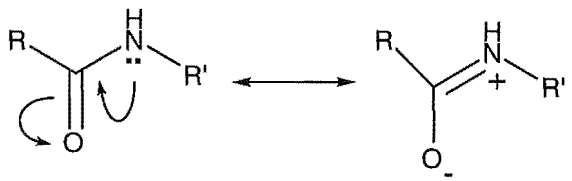
1.5.1 Introduction

The peptides are an important group of compounds. They are widely distributed in the human body, terrestrial plants, miroorganisms and marine organisms. Most of the peptides play a vital role in various life-processes, and many show remarkable biological activities. The peptides can act as chemical messengers, neurotransmitters, hormones, pain-killers, poisons, antibiotics, or food additives.[80] Among a variety of applications of the peptides, their utilisation for medical purposes has gained most importance. A well known example is penicillin (a modified tripeptide), a lifesaving antibiotic. Since penicillin was discovered in the 1940's, it has saved thousands and thousands of lives. Another peptide, angiotensin II, can cause an increase of blood pressure in the human body. A series of antagonists of this peptide have been synthesised to treat high blood pressure.

Normally, molecules composed of less than 50 amino acid residues are classified as peptides. Although the building blocks of the peptide molecules are relatively simple (the amino acids), the structures of peptides can be constructed in nearly countless ways due to the varied arrangements of the regular amino acids as well as the incorporation of abnormal amino acids and non-amino acids. The linkages between individual residues within peptides is the amide bond. A typical amide bond is shown below:



The amide bond does not exhibit the properties of the normal amino group or carbonyl group, as the nitrogen lone pair conjugates with the carbonyl group, influencing the bond lengths of C-O and C-N. The amide bond is best described by the following resonance forms:



The amide group is planar as the conjugation requires the nitrogen, carbon and oxygen p-orbitals all to lie in same plane. The peptide bond is relatively inert to most chemical reactions, with the carbonyl group only attacked under forcing conditions. The nitrogen does not usually exhibit nucleophilic tendencies.

As a prerequisite to completely understanding the biological properties or synthesis of the peptide, the structure of the peptide must be established. Normally, the natural peptides exist as a complex mixture in which the peptides are varied by a few residues, or by a change in the order of the residues. So, once pure the isolation and purification procedures are necessary to obtain a single congener for further study. The sequence of the constituent amino acids can be determined to establish the primary structure. Finally, the stereochemistry of the peptide can be addressed by various chemical and spectral methods.

1.5.2 Isolation and Purification of Peptides

Peptide molecules can be separated based on their size, charge at a given pH and polarity.

Like their component amino acids, the peptides possess an amino group at one end, a carboxylic group at the other end, and a series of sidechains separated by amide bonds. The peptide molecules can have a net charge at any pH, and the net charge is greatly affected by the nature of the side chains. The charged peptides can be absorbed by ion exchange resins, and their migration rate on the resins varies with pH and polarity of the eluant. Usually the peptides move on a column faster at their isoelectric point. One can take advantage of these properties to separate quite similar molecules. Another separation technique utilising the charge properties of a peptide is electrophoresis. In this method, a mixture of the peptides is loaded onto a solid support at a defined pH (either a sheet of filter paper, or a polyacrylamide gel). When a voltage potential is applied, the positively charged peptides move towards the cathode, while the negatively charged peptides migrate towards the anode. Alternatively, an isoelectric focusing technique can be used for difficult separation problems. In this technique, peptides which differ only slightly in their net charge at a given pH are required to migrate across a pH gradient. The longer the electrophoretogram is run, the sharper the bands become.

A peptide mixture can also be separated by gel chromatography based on either their size, or both size and polarity. The former principle is applied to gel filtration, and the latter to gel permeation. The large molecules can not migrate into the pores of the gel, so they are eluted first. In contrast, small molecules can diffuse into the pores and interact with the inner surface, thus causing selective absorption of polar

substances with low molecular weight.[81] The sizes of the peptide molecules in a mixture varies with the number and side chain type of the constituent amino acids. Gel filtration chromatography is suitable for mixtures containing a range of peptides of different molecular weight (size). However, it is very rare to obtain a pure peptide using this technique.[80] It is more practical to use gel filtration for the separation of peptides from proteins, polysaccharides and low molecular weight non-peptidic components. In the natural products field, the gel permeation technique, based on LH-20, is frequently used for separation of compounds in the molecular mass range <4000. Sephadex LH-20 has dual lipophilic and hydrophilic properties. Its lipophilic properties come from the isopropyl groups introduced by hydroxypropylation of Sephadex G-25, while the hydrophilicity is due to the numerous hydroxyl functions present. Depending on the polarity of the mobile phase, and the compounds to be separated, it is possible to perform normal phase, reversed phase and selective adsorption chromatography on LH-20.[82] The peptides may be separated based on both molecular size and polarity with a carefully selected single solvent, or solvent mixture. This is a very useful technique for elimination of other secondary metabolites to yield a 'pure' peptide mixture. Usually a pure peptide can be obtained subsequently by using high performance liquid chromatography (HPLC).

The excellent resolving power of the reversed-phase (RP) mode of HPLC has resulted in this technique becoming the predominant method for peptide separations,[83,84] although other modes are encountered in the literature. Solid phases for RPHPLC are mainly chemically-bonded phases of silica gel, for example, C-18, C-8, C-4, CN and phenyl. The fully end-capped C-18 or C-8 coated silica gel beads are widely used. However, a comparison study on different reversed-phase packings

revealed that the CN phase is very potent in separation of peptides and proteins.[84] The CN column showed excellent selectivity for peptides with varied hydrophobicity and peptide chain length. Peptides which could not be resolved easily on the C8 column were readily separated on a CN column.[84] River *et al.*[85] also pointed out the importance of the pore size of the beads, which played an important role in the chromatographic process. A pore size of 300Å *versus* 80-120Å was recommended for both good recovery and peak symmetry. Among various solvent systems, methanol-water and acetonitrile-water systems are most widely used. Incorporation of trifluoroacetic acid (TFA) in the mobile phase as a polar modifier can provide excellent conditions for the separation and purification of peptides. TFA has a significant effect on both peak shape and retention time in the concentration range 0.01-0.25%.[86]

1.5.3 Amino Acid Analysis

In order to identify the amino acids present, complete hydrolysis of the peptide to yield the free amino acids is necessary. The standard conditions include dissolving the peptide in 6M HCl and heating the acid solution at 110°C for 24 hours. However, caution must be taken because some amino acids can be destroyed, or modified under these conditions. For instance, tryptophan is completely destroyed during acid hydrolysis, amide side chains (eg., asparagine and glutamine) will be hydrolysed to the corresponding acids and threonine and serine can be slowly dehydrated to give corresponding alkenes. Some problems are also encountered with aliphatic amino acids, as there is steric hindrance due to the bulky side-chains. When Val-Val, Val-Ile, Ile-Val or Ile-Ile sequences exist in the peptide, careful interpretation of the data is necessary, because the cleavages of these bonds are incomplete under standard conditions.

The dansyl method (see Section 1.3.4) can help to identify such dipeptide sequences. In the case of uncertainty about the correct composition of a peptide, a second hydrolysis of the peptide for 72 hours should be performed. Methionine and carboxymethylcysteine are partially oxidised during the isolation of small quantities of peptides. The addition of 2-mercaptoethanol (0.05% (v/v)) will give better results for some amino acids.

Alternatively, the peptides can be hydrolysed with 2M NaOH (aq) at 100°C. Alkaline hydrolysis does not affect tryptophan, but will destroy serine, threonine, cystine and arginine. This method also causes racemisation of the amino acids. Thus, acid hydrolysis is the preferred method.[87]

Characterisation of the released amino acids can be achieved by chromatographic techniques. Of the various possibilities, such as paper chromatography (PC), thin-layer chromatography (TLC), gas chromatography (GC), high performance liquid chromatography (HPLC) and ion-exchange chromatography (IEC), the last technique has been the most widely used since the classic work of Moore and co-workers.[88-90] The technique was soon automated with slight modification, and routinely used in biochemistry laboratories. Today it is still one of the most reliable methods for amino acid analysis. In comparison with GC, however, the main drawback of the ion exchange technique is the low sensitivity and relatively long operating time. These drawbacks remained even when ninhydrin was replaced by fluorescamine,[91] pyridoxal,[92] or radioactive reagents[93] for detection.

The advantages of the GC method are obviously the low cost and the much greater versatility of the instrument compared with a specialised

amino acid analyser. When GC is coupled with mass spectrometry (GC/MS), it offers a more powerful method for identifying unknown amino acids, or non-amino acid components (see Section 4.3).

A disadvantage, however, is the necessity to derive the hydrolysate to produce volatile, less polar compounds that are suitable for GC analysis. Successful analysis of amino acids with GC is dependent on the synthesis of derivatives that are stable and reasonably volatile. Among the derivatization methods, the most preferred are those that require at least two derivatization steps, eg., esterification of the carboxyl group followed by acylation of the α-amino and remaining functional groups. A variety of alcohols have been used for the esterification step. These include methanol, propanol, isopropanol, butanol, isobutanol and isoamyl alcohol. The resultant amino acid esters are next acylated by the addition of acetic anhydride, trifluoroacetic anhydride (TFAA), pentafluoropropionic anhydride (PFPA), or heptafluorobutyric anhydrides (HFBA) along with an appropriate solvent, eg. CH_2Cl_2 or ethyl acetate. However, the only acyl derivatives that permit quantitative GC amino acid determinations are trifluoroacetic (TFA) and heptafluorobutyric (HFB), and with some restrictions, acetic anhydride.[94] Taking account of the difficulty of preparation, the resolution on the GC columns, and the losses of the more volatile derivatives during evaporation of the solvents, it is the N(O,S)-HFB isobutyl esters, the N(O,S)-HFB n-propyl esters, and N(O,S)-TFA n-butyl esters that have been used most extensively.[97]

In the early studies, amino acid analysis by GC was mainly performed on a conventional packed GC column. The polyester phases occupied a predominant position. Other types of stationary phases, such as polyglycols and silicones, were employed less as they were not able to

separate the isomers.[94] In the last decade, with the introduction of fused silica capillary columns and the development of immobilised stationary phases, the capillary column showed more potential in amino acid analysis.[96] The application of RPHPLC, utilising phenylisothiocyanate derivatives, has shown a dramatic increase recently.[97]

1.5.4 Sequence Determination

Sequencing is a key procedure for the structural analysis of the peptides. The choice of the method for peptide sequence analysis will govern to a large extent the whole approach to cleavage and separation of the resultant amino acids. The procedures include determination of the amino, or carboxy-terminal residues, and sequencing of the peptides by removal of one residue at a time.

For N-terminal determination, the dansyl chloride method and Dabitic method are frequently used due to their high sensitivity, without the requirement for expensive analytical equipment. The N-terminal amino acid of the peptides are reacted with dansyl chloride in $NaHCO_3$ solution. This is followed by acid hydrolysis (6M HCl). The hydrolysate is chromatographed along with a standard mixture of dansyl amino acids, on a polyamide layer plate. The dansyl amino acids can be identified by inspection of the spots under the UV lamp.[98] The Dabitic method uses 4-N,N-dimethylaminoazobenzene-4'-isothiocyanate as the coupling reagent. The advantages of this method over the dansyl method are that asparagine and glutamine are distinguished from aspartic and glutamic acids respectively, and that tryptophan may be determined. However, the determination of serine, threonine and lysine is less satisfactory. Alternatively, reverse-phase HPLC may be used for the detection.

The identification of the carboxy terminus is generally less easy than the N-terminus. One approach employs hydrazine, which reacts with peptides to cleave the amide bonds. Eventually, all of the constituent amino acids in the peptide are converted to hydrazides, except for the C-terminal residue, which will be present as the carboxylate. This can then be identified by amino acid analysis.

The most common method of sequencing peptides is the Edman degradation, which was developed by Edman in 1950.[99] The degradation is a cyclic procedure, by which the amino acid residues are cleaved one at a time from the N-terminus of the peptide and identified as phenylthiohydantoin derivatives. However, the overlap caused by incomplete coupling and cleavage reactions at each cycle together with a gradually increasing background resulting from small amounts of non-specific peptide bond cleavage during the degradation limit the application of the Edman degradation to, at the most, about 70 residues using automated sequencers, or about 30 residues using manual techniques. Thus, the polypeptides are normally chopped down with either enzyme cleavage, or partial acid hydrolysis.[100] After separation procedures, the individual peptide fragments are subjected to Edman analysis. The whole sequence can be established by the overlaps between these fragments. There are several other factors which may interfere the Edman method. For instance, blockage of the Edman degradation may occur when peptides contain asparagine, glutamine, tryptophan, S-carboxymethylcysteine and O-acetyl serine, when the N-terminus is blocked with formyl, acetate, methoxyacetate etc., when an abnormal amino acid (eg. β-alanine) or novel amino acids exist in the peptide molecules, or when the peptides possess a cyclic structure. The last three situations are frequently encountered in the study of peptides from micro-

organisms and marine organisms. The use of FAB mass spectrometry is a technique well suited to by-passing this sort of problem. This technique is not limited by N-terminal blocking, or the presence of rare residues, or even cyclic peptides.[101] Meanwhile, it provides very rich information about the structures of the amino acids. So, it is extremely useful when new amino acids or non-amino acids are present in the peptide chain (cf. Section 4.5). When the MS/MS mode is applied, it can even analyse peptide mixtures.[102] Recently, the development of an Edman method coupled with reversed-phase high performance chromatography (RPHPLC), together with the powerful computer-based quantitation and interpretation of chromatograms, has made automated sequencing extremely reliable.[103] Furthermore, with the development of 2D NMR techniques, such as 2D TOCSY, HMQC and HMBC, more and more sequencing work can be accomplished by NMR analysis (see Chapters 4-6 and related references).

1.5.5 Stereochemistry of Peptides

X-ray crystallography is the most straightforward method to study the structure and stereochemistry of the peptides. As soon as the relative configuration is established by X-ray analysis, the absolute stereochemistry of a peptide can be elucidated by determination of the configuration of one or two constituent amino acids. However, crystals suitable for X-ray analysis are not readily obtainable in most situations. The stereochemistry of peptides is usually first attempted using other techniques.

The typical techniques in the study of the stereochemistry of peptides include GC or HPLC analysis of either free amino acids, or their derivatives. The original procedures for resolution of the enantiomers of

amino acids involved the introduction of a chiral auxiliary to the carboxyl or amine functional groups of the amino acids. The resultant diastereoisomers could be resolved on optically inactive GC stationary phases.[104,105] However, since a chiral stationary phase for capillary column GC was developed by Gil-Av,[106] this type of stationary phase has become predominant in the studies of amino acid enantiomers. Of various chiral GC columns, Chirasil-Val, which consists of N-propional-L-valine tert-butylamide coupled to a co-polymer of carboxyalkylmethylsiloxane and dimethylsiloxane, is most widely used. A very good resolution can be achieved when the amino acid enantiomers are converted into N(O,S)-isopropyl esters. Recently, a rapid derivatization method was developed by Abe *et al.*, in which the amino acids were converted into N(O)-2,2,2-trifluoroethoxycarbonyl 2',2',2'-trifluoroethyl esters.[107] With this type of derivative, 18 pairs of protein amino acid enantiomers were resolved on a Chirasil-Val column. It is noteworthy that two pairs of enantiomers of isoleucine and *allo*-isoleucine were well separated in one run.

Another efficient technique is RPHPLC. The resolution of enantiomers can be accomplished for either the free amino acids on a chiral column, or by separating derivatives of the amino acids on a non-chiral column. Generally, the latter method gives better results. This latter approach can be achieved in two ways: one uses dansylated amino acids (Dns-Am) with a chiral additive to the mobile phase;[108] the other employs diastereomers produced from the reaction between amino acids and 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide using the normal HPLC solvent (Marfey's method).[109] Marfey's method has been successfully used in the study of several peptides from marine organisms. A new trend is the

coupling of HPLC with mass spectrometry (LC/MS), which provides more information for structural studies.

NMR techniques, particularly the NOE, NOESY and ROESY techniques, are very powerful in the study of conformations of the intact peptide. This has led to a better understanding of the relationship between the conformations and the biological activities of the peptides.[110]

1.5.6 Peptides from Marine Sources

The study of marine natural products started to boom in the 1970's, but the early literature was dominated by reports of non-nitrogenous metabolites.[111] It is only recently that marine peptides have become one of the most rapidly growing groups of marine natural products. Since Fusetani *et al.* isolated the first bioactive peptide, discodermin A, from marine sponge *Discodermia kiiensis*,[112] more than 400 peptides have been reported from marine organisms. It is interesting to note that 80% of them possess a cyclic structure. A high incidence of unusual amino acids or fatty acids have also been observed in marine peptides.[111] These marine peptides exhibit a wide range of bioactivities, such as cytotoxic, antimicrobial, antifungal, anti-viral and immunomodulating activity. This group of natural products has great potential in the search for new drugs and the development of pharmaceuticals.

1.6 Aims of This Project

As a part of the continuing search for bioactive natural products from marine organisms in the Marine Group of the University of Canterbury, this project includes:

A. Chemical screening study. This part of the work involved testing a modified protocol for chemical screening and searching for suitable organisms for further chemical study by screening a wide range of samples.

B. Investigation of anti-HIV components from a sponge and an ascidian. The primary tests indicated some HIV-inhibitory activity from a *Chondropsis kirkii* (sponge) and a *Styela* sp. (ascidian). Further separation was necessary to reveal the chemical properties and biological profile of the active compounds.

C. Studies of antiviral and cytotoxic constituents in a *Callyspongia* sp.

D. Separation and structural studies on the antitumour cyclic peptolides from *Lamellomorpha strongylata.* This sponge contains several classes of bioactive compounds. This project will focus on the peptolide fraction.

Chapter 2

Chemical Screening

2.1 Introduction to Chemical Screening

Marine natural products is a rapidly growing field. Over the last two decades, more than 7,000 compounds have been identified.[17] These discoveries have made a great contribution to both chemistry and biology. However, it has also caused problems for chemists because known compounds are being encountered with increasing frequency as the number of publications rise. This creates problems for chemists trying to find unique compounds for both publication and patenting. For those chemists using bioassay-guided separation techniques in searching for biological compounds, the presence of recurring compound classes is also a troublesome factor because it diverts the researcher's attention from other extracts with potential. Another problem in the day-to-day work of marine natural product chemists is that some of the bioactive components in the crude extract are sensitive to most separation media, especially as they are purified. To solve these problems the chemist requires an insight into the general properties of the interesting (bioactive) components before initiating a formal separation procedure on a large scale. A technique developed for this particular purpose has been called "chemical screening".

The concept of "chemical screening" has been in use for a while, but in the chemical sense the meaning has varied. In the work done by Grabley[113] and Noltemeyer *et al*[114] the term chemical screening meant screening for secondary metabolites by means of functional group- or compound-class-specific TLC detection reagents. Their methods were not related to the biological activity of the targeted organisms, but to chiral synthons for synthesis. These approaches turned out to be very effective

for the discovery of new natural products, and have been successfully used by many natural product chemists.

With a greater emphasis being placed on the discovery of the bioactive secondary metabolites, another chemical screening protocol was developed and successfully applied to the dereplication and prioritization of the crude aqueous extracts from terrestrial plants and lichens, cultured cyanobacteria, and marine invertebrates and algae as part of an anti-HIV study[115]. Later, the protocol was further modified by Blunt *et al*[116] and its application was extended to organic extracts with cytotoxic, anti-HIV, anti-viral, anti-bacterial and anti-fungal activities. It turned out to be a quick, simple and cheap method for optimising the isolation procedure for both aqueous and organic extracts. It is now a routine screening technique used by Marine Group at the University of Canterbury.

The general procedure adopted in this project includes the extraction of samples using suitable solvents, the rapid separation of the crude extracts with solid phase extraction cartridges, bioactivity testing of the eluates from the individual cartridges, interpretation of the bioassay results and decision making. Different types of solid phase extraction cartridges were carefully selected in order to obtain the information about polarity, neutral or ionic character, and molecular weight range. All cartridges were pre-equilibrated with an appropriate solvent. Then an aliquot of an extract (~5 mg) was either dissolved in a suitable solvent, or coated onto the celite before loading onto each cartridge. Generally, three fractions from each column were collected, but a fourth is an option. Each of the derived fractions were then tested along with a primary standard in the bioassay system.

For organic extracts, C18, CBA and LH-20 are frequently used, while the Diol, amino or PEI cartridges are chosen as alternative phases. Usually, MeOH/H_2O (1:1), MeOH and MeOH/CH_2Cl_2 (1:1) solvent systems are used for the C18 cartridge; MeOH/0.05M aqueous NH_4Ac (4:1), MeOH/2% aqueous ammonia (4:1) and MeOH/0.05% aqueous TFA (4:1) for CBA phase; while MeOH was used for LH-20. The C18 and CBA column can be run either under gravity or vacuum, but the LH-20 cartridge should only be operated under gravity.

Normally, the aqueous extracts are processed with C18, CBA and amino phases. Wide-pore C4 and Sephadex G-25 are also useful. The C18 cartridges are eluted with H_2O, MeOH/H_2O (1:1) and MeOH in turn, while the latter two cartridges are run with the following solvent systems: CBA, 0.05M of aqueous NH_4Ac, 2% aqueous ammonia and MeOH/0.05% aqueous TFA (1:1); amino, 0.05M of aqueous NH_4Ac, aqueous 0.05% TFA and MeOH/0.05% aqueous TFA (1:1) respectively.

The formal separation of the bioactive compounds on a large scale can then be initiated based on the interpretation of the screening results. The information from such a chemical screening exercise is illustrated by a typical model as below.

	C18	CBA	LH-20
I			+
II	+	+	
III			

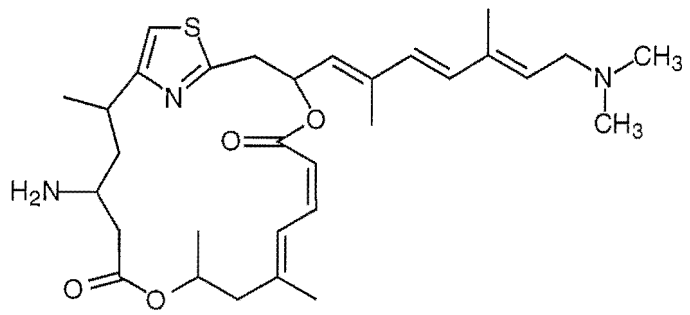
This profile implied that the bioactive compound(s) was of medium polarity, relatively large size (MW>750) and cationic. An optimised separation procedure can be readily established from the analysis.

When the above method is coupled with HPLC using photo-diode array (PDA) UV detection and a taxonomy-oriented database (MarinLit),[17] the approach becomes even more productive. The peaks in common in the HPLC traces of the bioactive fractions are likely candidates for the bioactive components. Comparison of the UV data for these peaks against that stored in MarinLit may result in the recognition of known structures or structural types. The current work has been carried out both to make some contribution to the modified chemical screening protocol, and to search for suitable candidates for further investigation.

2.2. Testing of the Modified Protocol

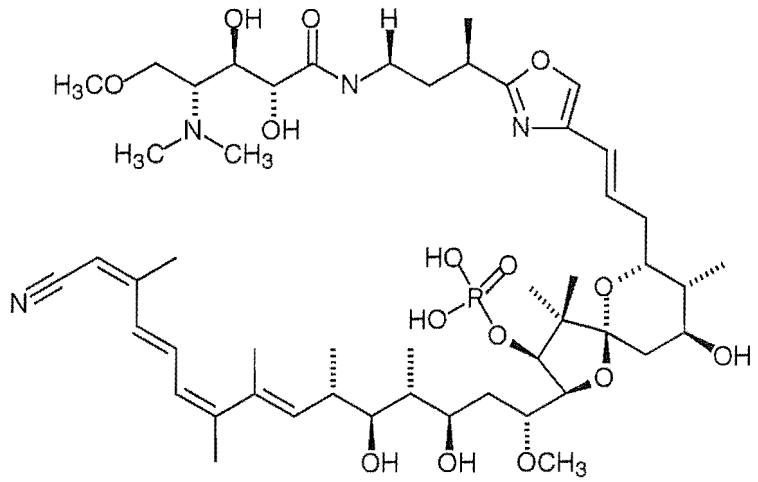
Like the development of any other method, the chemical screening approach has been improved step-by-step. A recent modification of the protocol included coupling of the results with HPLC and the MarinLit database. In order to test the efficacy and validity of the modified method, two sponges, *Mycale* sp.[117] and *Lamellomorpha strongylata*,[118] containing known bioactive compounds were chosen for detailed examination. They contain the cytotoxic compounds pateamine (**31**) and calyculin E (**32**) respectively.

An organic extract of the *Mycale* sp. was chromatographed on C-18, Diol, PEI and LH-20 cartridges using the procedure described in Section 2.1. The bioassay results indicated that the activity appeared in the second fraction of C18, the third fraction of Diol, the second fraction of PEI and the second fraction of LH-20.



Pateamine (**31**)

The structural conclusions from the data were medium polarity, cationic and medium sized. Then each of the active fractions was analysed by reverse phase HPLC, using an acetonitrile/H_2O gradient. The common peak (R_t=3 min) was observed in all HPLC traces. A UV spectrum corresponding to this peak indicated that the related compound(s) have



Calyculin E (**32**)

maximum absorptions at 274 and 285 nm, which was consistent with the UV data of pateamine. Obviously the active compound can be quickly located by the new screening procedure.

The organic extract of *L. strongylata* was passed through C-18, Diol, PEI, CBA and LH-20 cartridges respectively. Fractions 2 and 3 from C18, and all three collections from Diol showed strong cytotoxicity. These data can be interpreted as active compounds of different polarity or the compounds eluting over a wide range of polarities. It turned out to be the latter. The activity in the first and the second fractions from the PEI and CBA respectively suggested that the active compounds were cationic. The activity was observed in the first two fractions of LH-20, indicating large to medium sized molecules. The active fractions were each analysed by RPHPLC and the results summarised in Figure 2.2.1. Searching of MarinLit with the UV data (λ_{max} 321, 227 nm) of the common peaks (320 and 580 sec.) indicated that one of the active components probably had a calyculin type skeleton.[119] This assumption subsequently proved correct, and four known calyculins and two new compounds, calyculinamides, were identified.[118] It is noteworthy that further

processing of another fraction with end absorption only in the UV spectrum resulted in the identification of a different group of cytotoxic compounds. This was the theonellapeptolides. A detailed discussion about their isolation and structure study will be given in Chapters 4, 5 and 6.

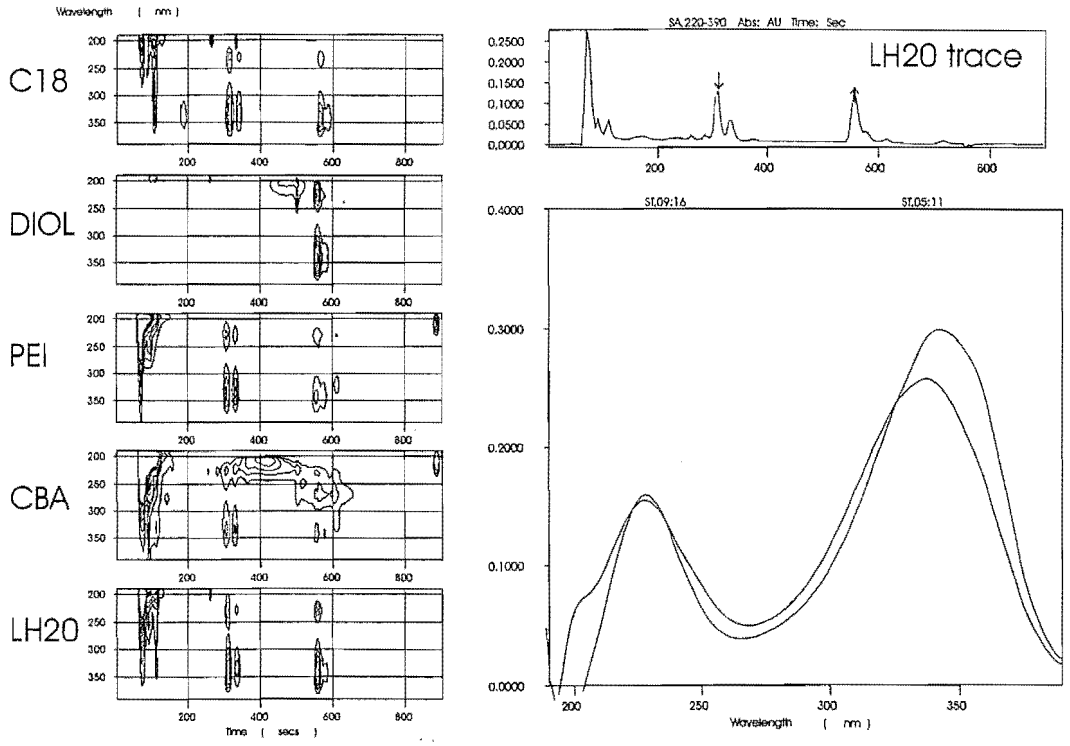


Fig. 2.2.1 HPLC analysis of the active fractions from *L. strongylata*. (a) Isoplots of the HPLC traces of the bioactive fractions from chemical screening (left); (b) normal HPLC presentation summed from λ 220-390 nm for the active LH-20 fraction(right top); (c) UV profile of the active components at 320 and 580 sec.

These experiments clearly demonstrated that the modified approach was very powerful for gaining insight into the properties of the bioactive components in the early stage of a study.

2.3 Screening for Anti-HIV Components from the Sponge *Chondropsis kirkii*

The primary bioactivity screening of the extract from *C. kirkii* showed medium to intense anti-HIV activity. But, there was no obvious cytotoxicity in P388 assay. The chemical screening was carried out using C4 wide pore, and Sephadex G-25 cartridges for the aqueous extract, and with C18, Diol and PEI phases for the organic extract.

Strong anti-HIV activity was found in the first fraction off the C4 cartridge, and in the first and second fractions off G-25. There was mild cytotoxicity in the former, but no cytotoxicity in the latter. This kind of chromatographic behaviour implied a high molecular weight for the active compounds. On the other hand, a medium anti-HIV (51-54%, inhibition) activity accompanied by mild cytotoxicity (31-34%, cytotoxicity) eluted in the third fraction off the C18 cartridge, and the first fraction off the Diol column. These data were interpreted as another group of less polar active components. No activity was observed in any fractions off the PEI column which suggested that the active molecules were polyanionic.

Because recurring compound classes are frequently responsible for the anti-HIV activity of marine aqueous extracts,[120] it was necessary to examine and dereplicate for these classes of compounds before embarking on a major effort in this project.

Sulphated polysaccharides are often responsible for the anti-HIV activity of aqueous fractions. It is necessary to remove this type of component in order to focus on other classes of active compounds. A 50% EtOH in water solution can precipitate the polysaccharides and protein, and reduce

interference from these compounds. Albano and Mourao[121,122] demonstrated the cationic dye toluidine blue bonded to polyanionic molecules in aqueous solution and induced a metachromatic shift in wavelength. This resulted in a decrease in absorbance with increasing concentration of the polyanionic molecules. They successfully applied this method to the determination of sulphated polysaccharides in sea cucumbers and tunicates. To find out whether sulphated polysaccharides existed in the supernatant of the aqueous extract of *C. kirkii*, the solution was run through a Sephadex G-25 column. Each fraction was analysed for sulphated polysaccharides at three different concentrations (see Experimental, Table 7.2.1). A decrease in absorbance with increasing sample concentration was observed for the two active fractions, but there was no obvious absorption change with concentration for the inactive fractions. This suggested that sulphated polysaccharides were present in the supernatant. However, Beutler *et al*[120] had noticed that colour interference from the crude extracts gave erroneous results in a significant number of cases. To obtain more evidence, the ethanol precipitate was redispersed into water and analysed in same way as for the supernatant after filtration. All seven fractions exhibited strong anti-HIV activity. The toluidine blue test indicated a positive response for the presence sulphated polysaccharides to all seven fractions as a function of different concentrations (cf. Table 7.2.1). These data gave more supporting evidences for the above assumption. The activity in the supernatant can probably therefore be attributed to the presence of trace amount of sulphated polysaccharides. Because this class of compounds was not a target of this project, no further study on them was carried out.

Chemical screening also indicated that the organic extract contained anti-HIV activity. The organic extract from *Chondropsis sp.* was loaded onto

a C-18 column, which was eluted with an increasing proportion of MeOH in water, MeOH/CH_2Cl_2 and CH_2Cl_2. The anti-HIV activity was found in the fractions eluted with MeOH and MeOH/CH_2Cl_2 (1:1). All the active fractions were combined on the basis of their TLC behaviour, and re-chromatographed on a Diol column. The anti-HIV activity was concentrated in just three fractions, accompanied by weak cytotoxicity. After LH-20 chromatography, the individual fractions were examined by ^1H, or ^{13}C NMR spectroscopy. The spectra indicated that the components were a sterol mixture. Reverse phase HPLC analysis indicated at least six partially resolved peaks, although only one spot was observed by TLC. An attempt to purify the steroid fraction by HPLC (CN column in a reverse phase mode) did not result in any pure compounds. Bioassay results on the fractions from the HPLC chromatography indicated that the level of anti-HIV activity had not varied much during the purification, but that the cytotoxicity had increased most notably (see Table 2.3.1). The HPLC of the most active fraction showed that this was still a mixture of up to six compounds, but with one compound predominant. The ^1H NMR spectrum (Fig. 2.3.1) of this fraction confirmed that one sterol was predominant, but difficulties were encountered in the interpretation of the data. The multiplet at δ 3.54 was typical of an H3 resonance for a sterol. A doublet at δ 4.68 (2H, J=15 Hz) was assigned to two olefinic protons with *trans*-geometry, and must therefore be located in side chain. However, a doublet at δ 5.38 (J=3 Hz) normally attributable to the H6 resonance of a 5(6)-double bond corresponded to two, rather than one proton, suggesting the presence of a third trisubstituted olelfin in a very similar chemical and magnetic environment to the 5(6) olefin. The ^{13}C NMR spectrum showed six unsaturated carbon signals between δ 121.7 and 140.7. Four methine carbons were recognised by the DEPT

experiment. The resonances at δ 140.7 and 121.7 were assigned to C5 and C6 respectively. These data implied that the major sterol had one hydroxyl group, and three double bonds in its structure. The acetylated derivatives of the sterols were also prepared and subjected to analysis on a variety of HPLC columns. No significant improvement was obtained in this way. As the samples were cytotoxic steroid in character and not strong inhibitors of HIV, it was decided not to proceed further with this sample.

Table 2.3.1. Anti-HIV activity and cytotoxicity of the fractions from *Chondropsis kirkii*

Sample Code	Antiviral % Inhibition	CEM Cells% cytotoxicity
Crude Fraction	89	21
SL235.4	59	49
SL236.1	50	22
SL236.3	47	26
SL240.7	63	45
SL241.2	72	44
SL242.3	87	47

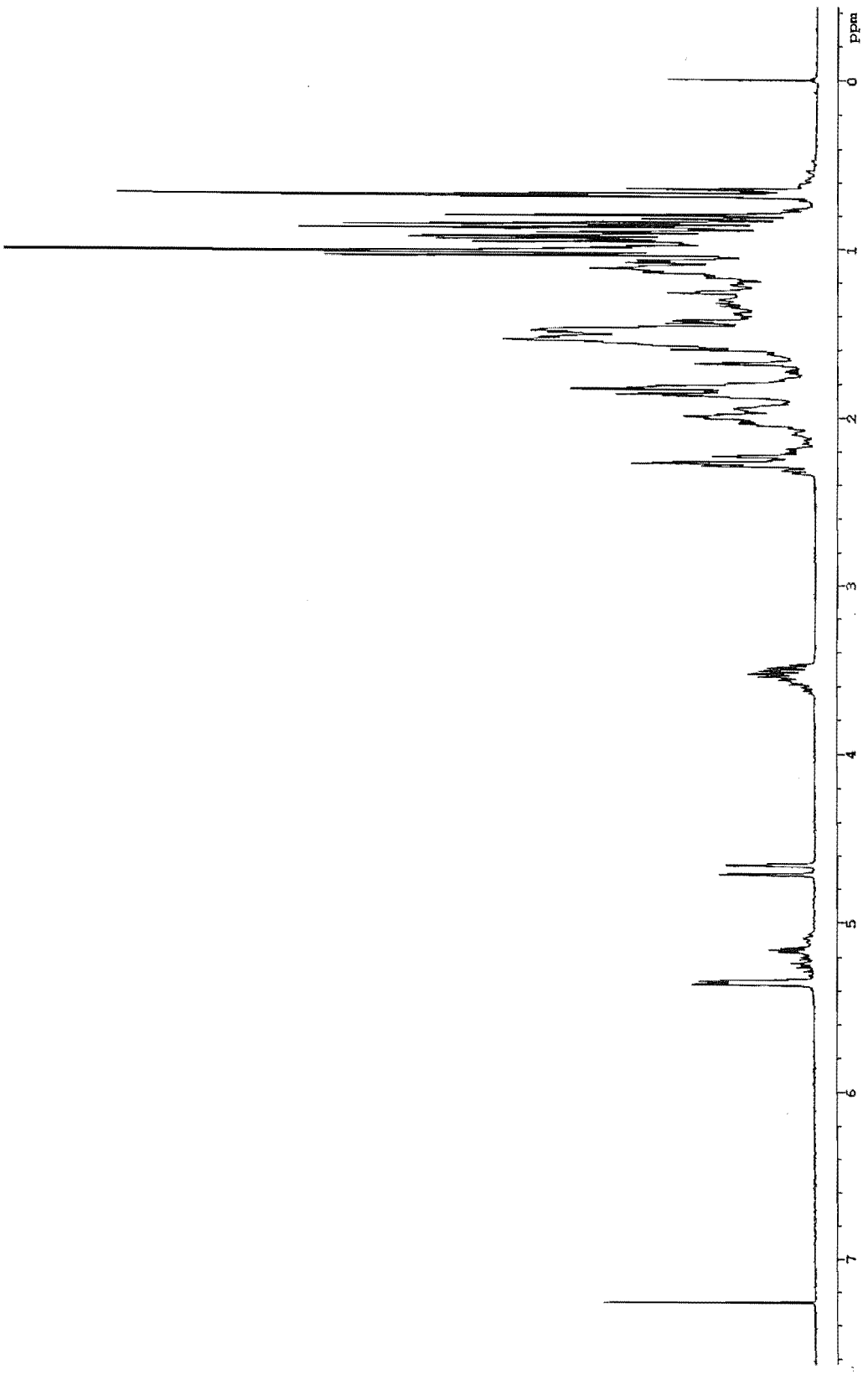


Fig. 2.3.1 ^{1}H NMR spectrum of SL242.3

2.4 Extraction and Chemical Screening of the Ascidian *Styela* sp.

The ascidian *Styela* sp. was extracted with MeOH, MeOH/CH_2Cl_2 (1:1) and CH_2Cl_2 in turn, followed by partition of the extracts between water and ethyl acetate to yield aqueous and organic extracts. An equal volume of EtOH was added to the aqueous fraction which was then filtered through celite to remove the precipitate and give a supernatant solution. The supernatant and organic extracts were each tested for anti-HIV activity. The former exhibited a 28% inhibition against the infected cells, while the latter only showed cytotoxicity in the assay. To further evaluate the anti-HIV activity, the aqueous extract was screened with C4 wide pore and the Sephadex G-25 cartridges. The second fraction from the C-4 cartridge showed good HIV-inhibitory activity (106% inhibition), and the first and fourth fractions from the same column had a weak activity. All four fractions from Sephadex G-25 cartridge were negative. Since the separation mechanism of G-25 is molecular size selection, the retained molecules must be of low molecular weight. Absorption of the bioactive material was also a possibility. The anti-HIV active compounds were considered most likely to be polar molecules with low molecular weight. Furthermore, their separation should rely mainly on reverse phase chromatography.

The aqueous extract was chromatographed on a C-18 column, eluted initially with water, then different proportions of MeOH in H_2O up to 100% MeOH. The first three fractions showed weak HIV-inhibitory activity. The signals between δ 7.9-8.8 in ^1H NMR spectrum (Fig. 2.4.1) of fraction 1 are related to either the amide protons of peptides, or the aromatic protons of nucleotides. The resonances between δ 4.0-5.0 could

be accounted for by the α-protons of amino acids, while the peaks between δ 3.1 and 3.9 suggested the presence of a sugar moiety. So, the existence of small peptides, or nucleotides in the active fraction were deduced. The above analysis also excluded any activity related to sulphated polysaccharides.

For the organic extract, C18, Diol and PEI cartridges were used for chemical screening. The cytotoxic activity was found in the second and third fractions off C18, the first and second fractions off Diol, and the third fraction off PEI. This chromatographic behaviour suggested that the active compounds were of medium-polarity to non-polar, and probably anionic.

Unfortunately, the limited amount of sample prevented further processing. Despite repeated attempts, it proved impossible to recollect this animal. The ascidian *Styela* sp remains a candidate for further study.

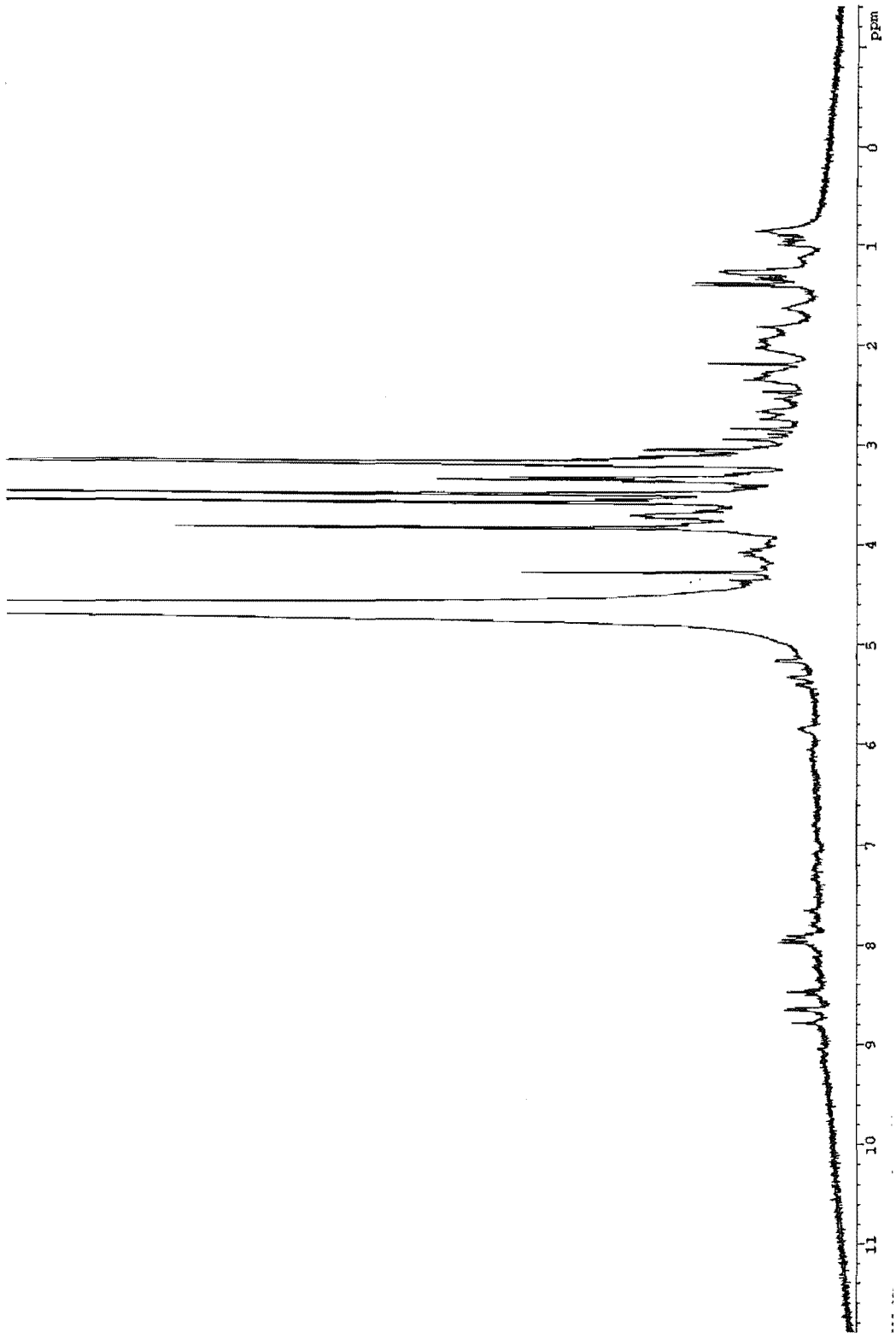


Fig. 2.4.1 ^{1}H NMR spectrum of fraction 1 from C18 column (*Styela* sp.)

2.5 Chemical Screening of an Unidentified Sponge (Code No. 94FL01-01)

Weak anti-HIV activity was observed in the crude extract from this unidentified sponge species. In order to obtain more information about the active compound(s), the frozen sponge (550g) was extracted by the same procedure as used for *C. kirkii.* (see experimental). Both the aqueous and organic extracts were analysed by the chemical screening protocol. The resultant fractions were sent to the USA for anti-HIV bioassay. The bioassay results revealed no remarkable anti-HIV activity in any fraction. This is a indication of the unstable bioactive components. No further work was carried out on this sample.

2.6 Chemical Screening for Anti-viral Activity of the Extract from *Callyspongia irregularis*

The primary bioassay for the organic extract from *C. irregularis* revealed very potent anti-viral activity with weak cytotoxicity (inhibition zones were 4+ against both *Herpes simplex* and *Polio* virus, while zone of cytotoxicity was +). Chemical screening of the organic extract was on C18, Diol and LH-20 phases. The anti-viral activity was found in the second fraction from the C18 column and in all three fractions from Diol, while only weak activity was observed from the last fraction from the LH-20 column. These data implied that the active components have medium polarity accompanied with some less polar compound, or there was strong tailing on the Diol phase, and medium to low molecular weight. The processing of this sample will be discussed in Chapter 3.

2.7 Chemical Screening of an Unidentified Sponge (Code No. 95CR4-18)

The organic extract from the unidentified sponge 95CR4-18 was screened on C18, Diol, PEI, CBA and LH-20 phases. Fraction 3 from C18, fraction 3 from Diol and fraction 2 from LH-20 exhibited potent cytotoxicity, while fractions from PEI and CBA showed only very weak activity. These observations suggested that the active compound(s) were polar to medium polarity. Meanwhile, a decrease in the bioactivity in all fractions, except those from LH-20, was noticed. The dilution, which is equivalent to the IC_{50} for the two most active fractions from the C18 and Diol phases, had risen to 526 and 233. This compared with the initial dilution of <97.5 for the fraction from LH-20 and standard. This resulted from either an experimental error, or by loss of the bioactive compound(s). The most likely scenario is loss of bioactivity on the C18 and Diol phases.

To find out the effect of the solid phases on the bioactivity, an active fraction was run through a C18 column. All collections were then tested for cytotoxicity. This time the most active fraction, corresponding to a MeOH eluate, showed a dilution value of 5541. An obvious decay in the bioactivity had been observed. In order to gain insight into the nature of the bioactive compounds, the HPLC with PDA detector was used for further study. Unfortunately, no common peaks among the various active fractions could be recognised. This was attributed to either UV inactive components, or a very low concentration of the active compounds. No further work was undertaken on this sample due to the instability of the active component(s).

2.8 Chemical Screening of an Unidentified Sponge (Code No. 95CR2-10)

An aqueous extract of this sponge exhibited very potent cytotoxicity (IC_{50}<97.5). The chemical screening was carried out using C4 wide pore, PEI, CBA and G-25 cartridges. Bioactivity testing of all 11 fractions revealed that only the first fraction from the CBA column was active, and then only weakly cytotoxic. The activity nearly disappeared after chromatography. However, the standard solution of the extract remained active (dilution <97.5). It was assumed that the active components were sensitive to all the normally used solid phases.

2.9 Conclusion

In the current work, a total of eight marine organisms were processed in a search for suitable candidates for further study. The results indicated that chemical screening is a very efficient tool for both dereplication and optimisation purposes in studying bioactive marine natural products. At the same time, the above experiments also explored limitations related to the approach. When the bioactive components exist in trace amount, or are UV inactive, the interpretation of the results is not straightforward. Of course, these shortfalls can be supplemented by "secondary chemical screening"[123] and coupling this protocol with other analytical means, for example, LC/MS or IR spectrometry. As a direct result from the current chemical screening, the sponges *Callyspongia irregularis* and *Lamellomorpha strongylata* were chosen for further chemical investigation. This work resulted in the identification of several anti-viral and cytotoxic compounds (see Chapters 3, 4, 5 and 6).

Chapter 3

Antiviral and Cytotoxic Compounds from Callyspongia *sp.*

3.1 Introduction

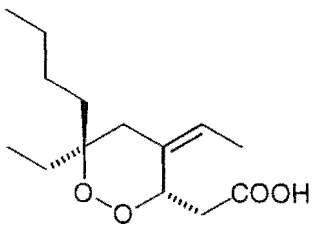
Callyspongia irregularis was chosen for further study because both very promising antiviral activity (HSV 4+ and PV 4+) and cytotoxicity against P388 cell line (177 x 10^{-6} dilution) had been observed in the primary bioassay. However, the NMR spectra of the active fractions from chemical screening indicated that the active components probably existed in very low concentration. Considering the available amount of the sponge sample (380 g of fresh animal), great care needed be taken in the separation procedure to minimise the loss of the active compounds. With this in mind, high speed countercurrent chromatography (HSCCC), and gel permeation techniques were adopted as the main separation means, while Diol column chromatography and RPHPLC were reserved for the final purification stages. By this strategy, a trace amount of the known, strongly antiviral compound, mycalamide A,[124] was identified. At the same time, several other cytotoxic fractions were located. In order to further explore the biological activity of this sponge, a recollection was carried out in 1995. In fact, another species of *Callyspongia* with similar characteristics to *C. irregularis* was collected. This species was named *Callyspongia* sp.2. Multistep bioassay-directed separation of the organic extract from *C.* sp.2 did not result in the isolation of mycalamide compounds, but rather two very cytotoxic fractions with IC_{50} 11 ng/mL and <97.5 ng/mL, were obtained.

3.2 Compounds Isolated from Genus *Callyspongia*

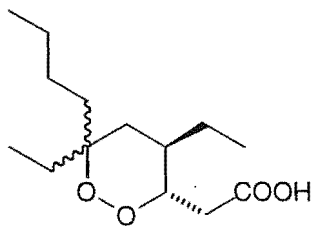
A number of compounds with a variety of bioactivities have been isolated from species of sponges from the genus *Callyspongia*. To date, not many species from this genus have been targeted by chemists. The typical biological activities of the components found in this genus include cytotoxicity,[125,126] inhibition of the epidermal growth factor (EGF),[127] inhibition of the fertilisation of starfish gametes, inhibition of phosphatidylinositol-specific phospholipase C (PI-PLC),[128] and anti-microfouling effect.[129] A feature of earlier work in this genus has been the isolation problems encountered.[127]

3.2.1 Cytotoxic Compounds from *Callyspongia* sp.

A number of straight-chain and branched fatty acids containing cyclic peroxides, and steroidal peroxides have been isolated from sponges of this genus. Many of the cyclic-peroxide-containing acids exhibited cytotoxic or other types of biological activity. Two branched-chain acidic compounds, (**33**) and (**34**), containing a cyclic peroxide moiety were isolated from a *Callyspongia* sp.[125] Their structures were elucidated by spectral methods. The peroxide partial structure was deduced from unsaturation elements in cooperation with NMR analysis. Acids (**33**) and



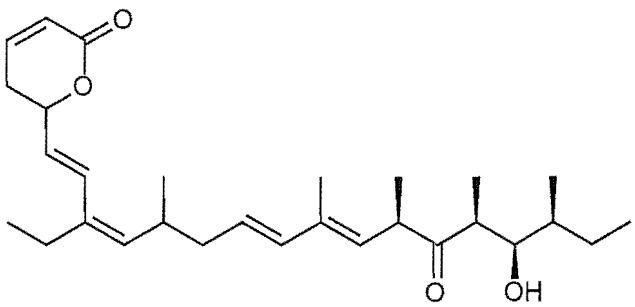
Peroxide (**33**)



Peroxide (**34**)

(**34**) inhibited murine leukaemia cell growth (ED_{50} (P388) 5.5 and 2.6 μg/mL respectively.)

Another cytotoxic compound, callystatin A (**35**), was identified from *Callyspongia truncata.*[126] It was characterised as a polyketide with a terminal α,β-unsaturated δ-lactone and exhibited cytotoxicity against KB cells (IC_{50} 0.01 ng/mL). The absolute stereochemistry at C-19 in callystatin A (**35**) was established by Mosher's method[130]. Since **35** was also found in another marine sponge, *Stelleta* sp, and is related to the antitumour antibiotics, eg., leptomycin,[127] kazusamycin[128] and leptofuranin,[129] isolated from the cultured broth of several strains of *Streptomyces* sp., it was hypothesised that symbiotic microorganisms participated in the biosynthesis of callystatin A.

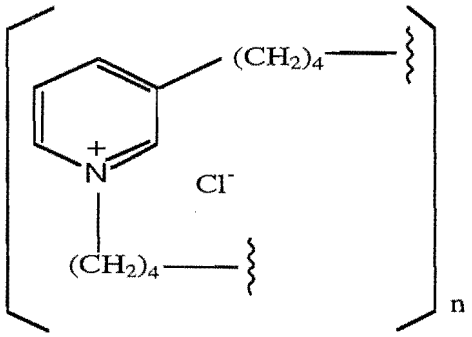


Callystatin A (**35**)

3.2.2 Other Bioactive Compounds from *Callyspongia* sp.

Davies-Coleman *et al.* reported isolation of EGF-active polymeric pyridinium alkaloid from the sponge *Callyspongia fibrosa.*[131] This compound is illustrative of the complexity and difficult factors sometimes encountered in the structural elucidation of polymeric compounds. The structure of the EGF-active compound was first proposed, on the basis of ion-spray mass spectrometry, to be the cyclic dimer **36** (n=2), but further comparison with synthesised **36** and its derivatives revealed that

hypothesis was incorrect. The mass spectrum from ion-spray mass spectrometry showed a predominant peak at *m/z* 379, which was interpreted as arising from a dimer structure (**36**). The molecular formula of the salt of **36** was therefore thought to be $C_{26}H_{40}N_2Cl_2$.

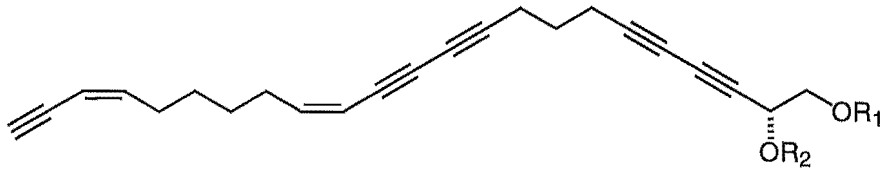


Dimeric pyridinium alkaloid (**36**) (n=2)

However, when **36** and several other derivatives were synthesised, and compared with the authentic natural product, the differences were obvious. Firstly, the natural product exhibited a different *Rf* value from the synthesised compounds by TLC. Secondly, the larger the molecule, the less abundant the parent ion observed in the ion-spray mass spectrum, while the natural product did not show the (M^{2+} + A^-) peak in its MS spectrum. Finally, the bioactivity of the synthesised compounds did not match that observed for the natural product. The difference in potency was at least 17-fold. Thus, the authors suggested the active natural product was a high molecular weight oligomer or polymer.

Several polyacetylene compounds have been isolated from *Callyspongia* species, but only two polyacetylene sulfates, callyspongins A (**37**) and B (**38**),[132] were reported showing starfish gamete fertilisation-inhibitory activity. In this experiment, the oocytes were induced to maturity by treatment with 1 μM 1-methyladenine. Maturing oocytes were fertilised 40 min after the start of the 1-methyladenine treatment. The oocytes

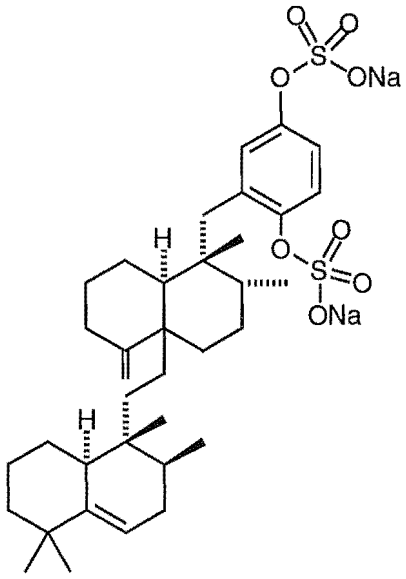
were examined for the presence of the germinal vesicle and the fertilisation envelope at 5 min after insemination. Compounds **37** and **38** inhibited the fertilisation of starfish gametes at minimum inhibitory concentrations of 6.3 and 50 μM respectively.



Callyspongin A (**37**), R_1=R_2=SO_3Na

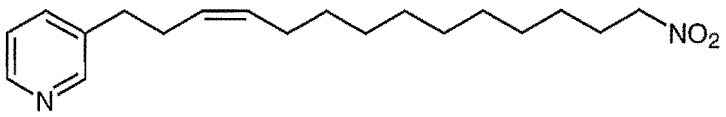
Callyspongin B (**38**), R_1=SO_3Na, R_2=H

Fukami *et al*[133] isolated an enzyme inhibitor, akaterpin (**39**) from an unidentified *Callyspongia* species. The structure was characterised by spectral methods. This new triterpene exhibited inhibitory activity against phosphatidylinositol-specific phospholipase C (PI-PLC) with an IC_{50} of 0.5 μg/mL. It also inhibited neutral sphingomyelinase with an IC_{50} of 30 μg/mL. Akaterpin was also a potent PI-PLC inhibitor.

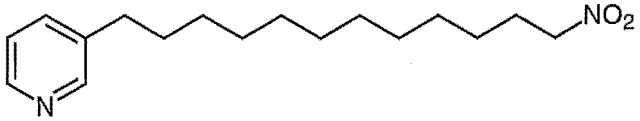


Akaterpin (**39**)

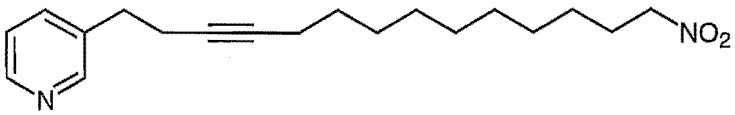
Wang *et al*[134] found three anti-microfouling nitroalkyl pyridine alkaloids, untenines A (**40**), B (**41**) and C (**42**), from the Okinawan sponge, *Callyspongia* sp. This was the first isolation of a nitroalkyl metabolite from a marine organism. Compounds **40**, **41** and **42** showed anti-microfouling activity with IC_{100} values of 3.0, 6.1 and 5.8 $\mu g/cm^2$ respectively. It is suggested that these rare nitroalkyl metabolites are the precursors of nitroso, hydroxylamine and amine derivatives, considered to be transient intermediates from nitro groups reduced by reductase in the presence of NADPH. Niphatynes[135] isolated from sponge, *Niphates* sp., provide more evidence for the assumption.



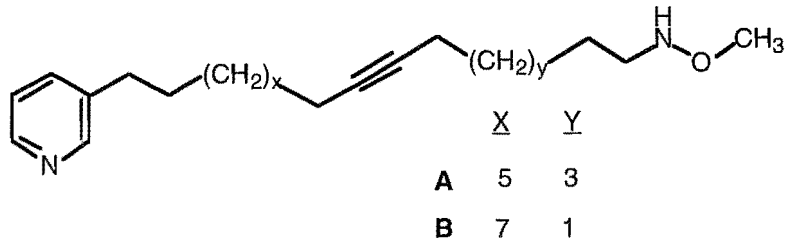
Untenine A (**40**)



Untenine B (**41**)



Untenine C (**42**)



Niphatynes A and B

3.3 Studies on Anti-viral Components from *Callyspongia irregularis*

C. irregularis possesses a tubular body form with a purple colour in sea water. The animals used in this study were collected at Moeraki (South Island, New Zealand, 15 m depth) by SCUBA.

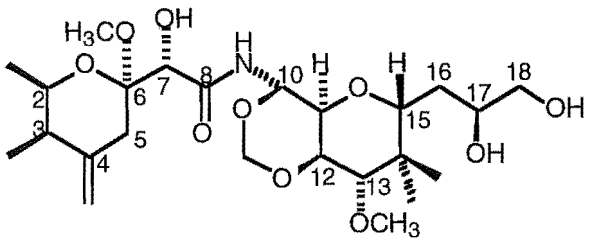
3.3.1 Extraction and Separation of Compounds from *C. irregularis*

The frozen sponge was blended and extracted with MeOH, MeOH/CH_2Cl_2 and H_2O respectively. After precipitation of the higher molecular weight species with ethanol, the aqueous portion was freeze-dried to produce an aqueous extract. This extract did not show any activity in a cytotoxicity test. However, the organic extract exhibited strong activity against *Herpes simplex* and *Polio* virus, and cytotoxicity in both the anti-viral bioassay and P388 assay.

The organic extract was chromatographed on C-18 eluted initially with H_2O, followed by increasing concentration of MeOH in water, MeOH and MeOH/CH_2Cl_2. The main activity was found in the fractions that eluted with MeOH/H_2O (7:3). The fractions from MeOH/CH_2Cl_2 (1:1) and CH_2Cl_2 also gave a mild cytotoxicity response. From these results it was concluded that the major active compound was polar , but that the less polar fractions also contained some activity. The eluant was combined into six fractions according to the bioactivity, and with reference to TLC characteristics. A further purification of the most active fraction was effected by high-speed countercurrent chromatography (HSCCC). The solvent system, petroleum ether/ ethyl acetate/ methanol/water (6:11:6:4) with the upper phase as the mobile

phase, was employed. This procedure successfully concentrated the bioactivity into just two fractions. The ^{1}H NMR of the active fractions displayed many resonances in δ 3.0-4.5 region. This suggested that a polyhydroxy or polyether compound was present in the mixture. The two active fractions were combined and the separation by repeated HSCCC. The resultant fractions were combined into five portions based on their biological activity. Of these, II.2 displayed significant anti-viral and cytotoxic activity, while II.5 showed only cytotoxicity (see Scheme 3.3.1).

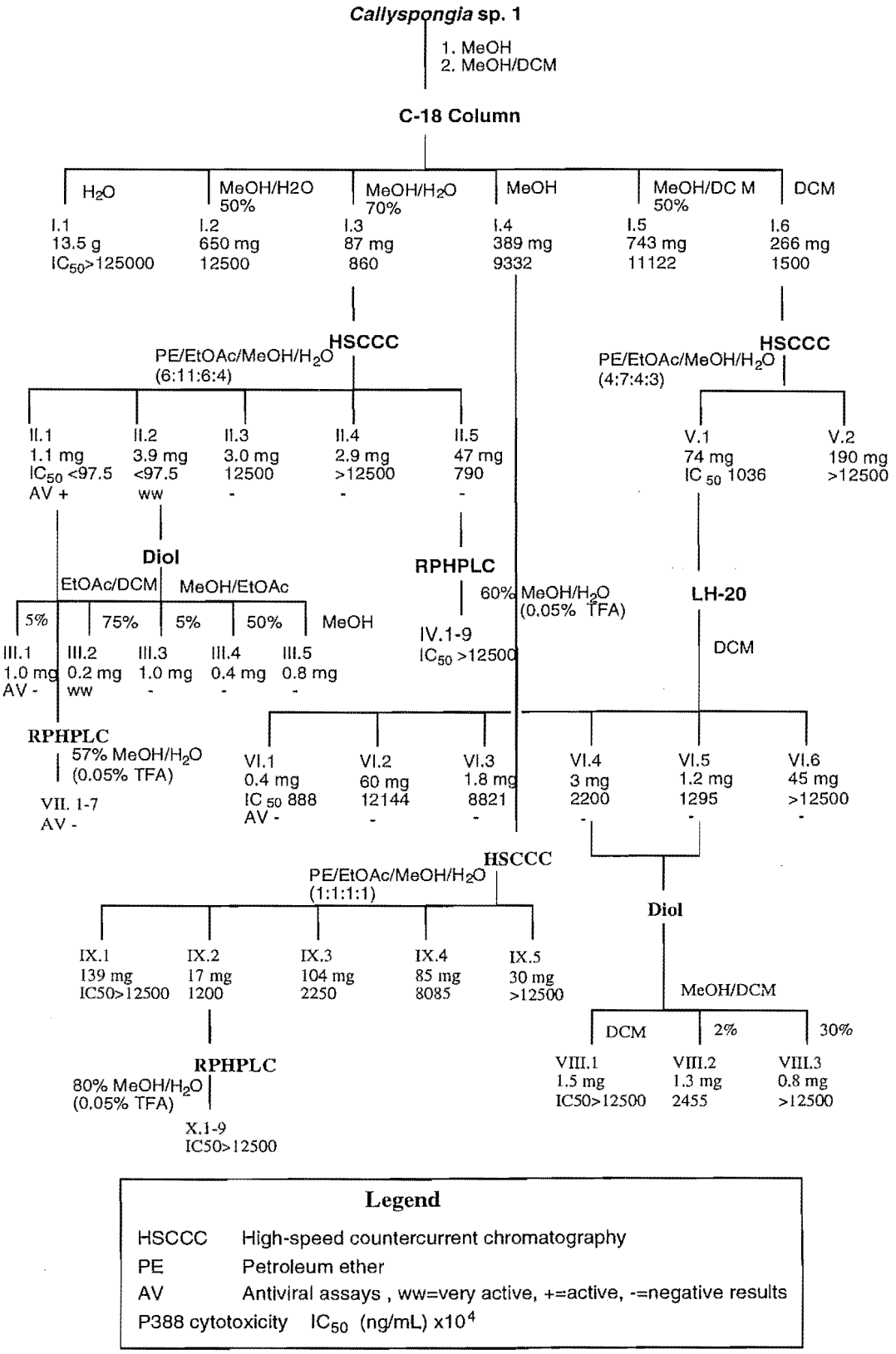
The anti-viral fraction was chromatographed on a Diol column. A compound, which showed whole-well inhibition against both *Herpes simplex* virus type I (a DNA virus) and *Poliovirus* type I (an RNA virus) at a concentration of 0.01 μg/mL, was isolated. The bioactivity profile of this compound was very similar to that of the antiviral compound, mycalamide A (**43**), previously isolated by the Marine group from a *Mycale* sponge.[124] After a careful examination of the spectral data for both the active compound and those of authentic mycalamide A, it was concluded that the compound isolated from *C. irregularis* was mycalamide A. The mycalamide A was present at a very low concentration in the sponge. Only about 200 μg was obtained in this work.



Mycalamide A (**43**)

To accumulate more sample for further characterisation, the less active, but related fraction II.1, was purified by RPHPLC using MeOH/H_2O containing 0.05% TFA (trifluoroacetic acid) as eluent. However, neither the antiviral, nor the antitumour activity was observed in any fractions from HPLC. In a similar way, a cytotoxic fraction (II.5) was submitted to RPHPLC for purification. Again, no bioactivity was found in any collections from the column. At this point a re-examination of the literature for the chemical properties of mycalamide compounds was made. In his thesis work,* Dr. Andrew M. Thompson had pointed out that mycalamide samples had no detectable activity after standing in 10% MeOH/aqueous hydrochloric acid (0.1M) for three days. So, it is most likely that the presence of the trace amount of TFA in the HPLC solvent contributed to the decomposition of mycalamide A during concentration and loss of the bioactivity.

* Thompson, A. M. *Synthetic Studies on a Potential Antitumour compound* (PhD. thesis), **1991**, pp.145-146.



Scheme 3.3.1 Bioassay-guided separation of the extract from *C. irregularis*

3.3.2 Identification of Mycalamide A

Since mycalamide A displayed significant *in vivo* anti-viral and antitumour activity, it has attracted wide attention from both chemists and biologists. Before the separation of mycalamide A from *C. irregularis*, it had been obtained only from a species of the genus *Mycale* (family Mycalidae Lundbeck, order Poecilosclerida).[124] This new discovery offers an additional source as well as raises very interesting questions as to the origins of the mycalamides. Is it a sponge metabolite, or it formed by a symbiont?

As a limited amount of the sample was available, the structural information could only be extracted from the ^{1}H-NMR spectrum and by MS analysis. In order to obtain more evidence to confirm the structure of the active compound, the sample was sent to the National Institutes of Health for high resolution mass spectroscopy (HRMS). The molecular ion could not be detected due to its very low intensity. Fortunately, a major ion at *m/z* 472, corresponding to the loss of CH_3OH from the molecular ion at *m/z* 503, was recognised both in the FABMS spectrum of the sample and the standard mycalamide A. An ion at *m/z* 472.2528 (Δ -1.9) in the former and *m/z* 472.2533 (Δ -1.3) in the latter were observed respectively, which correspond to a molecular formula of $C_{23}H_{38}O_9N$. A comparison of the link scan from *m/z* 503 for the *Callyspongia* sample with that obtained for mycalamide A revealed that both of them yielded a reasonably intense ion at *m/z* 472, and that nearly all other major daughter ions from *m/z* 503 are the same, although the ratios of the daughter ions varied between the two spectra (Fig. 3.3.1 and 3.3.2). This deviation can be attributed to the co-existence of impurities in the *Callyspongia* sample, that affected ionisation conditions. The major ions,

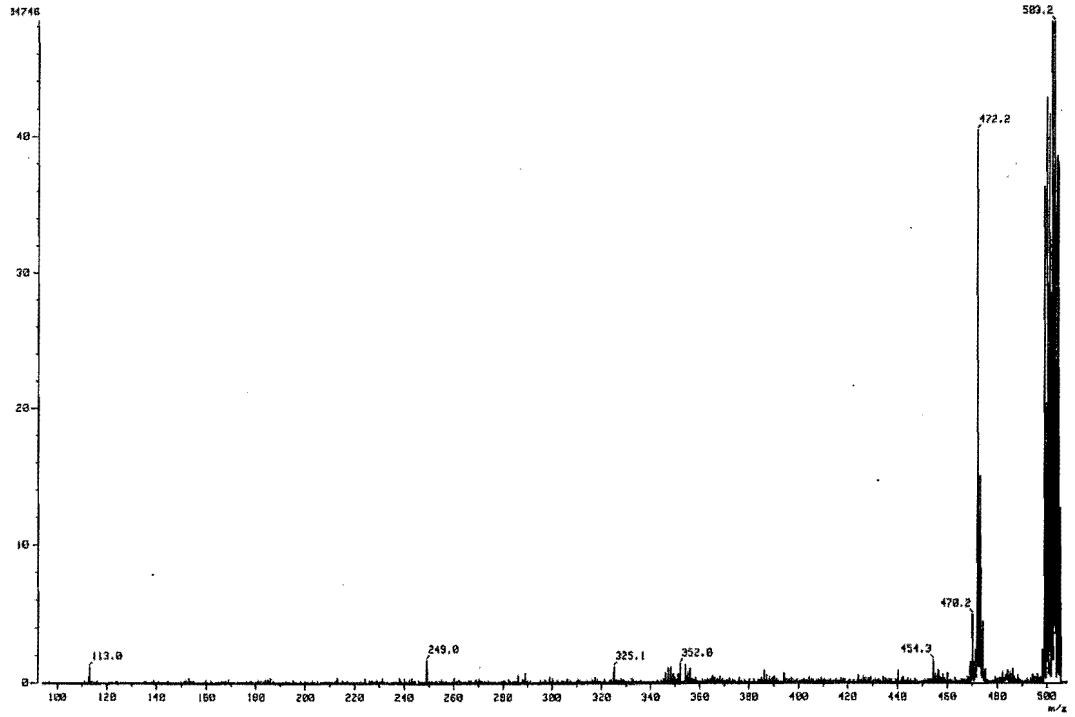


Fig. 3.3.1 Link Scan at *m/z* 502 for Mycalamide A

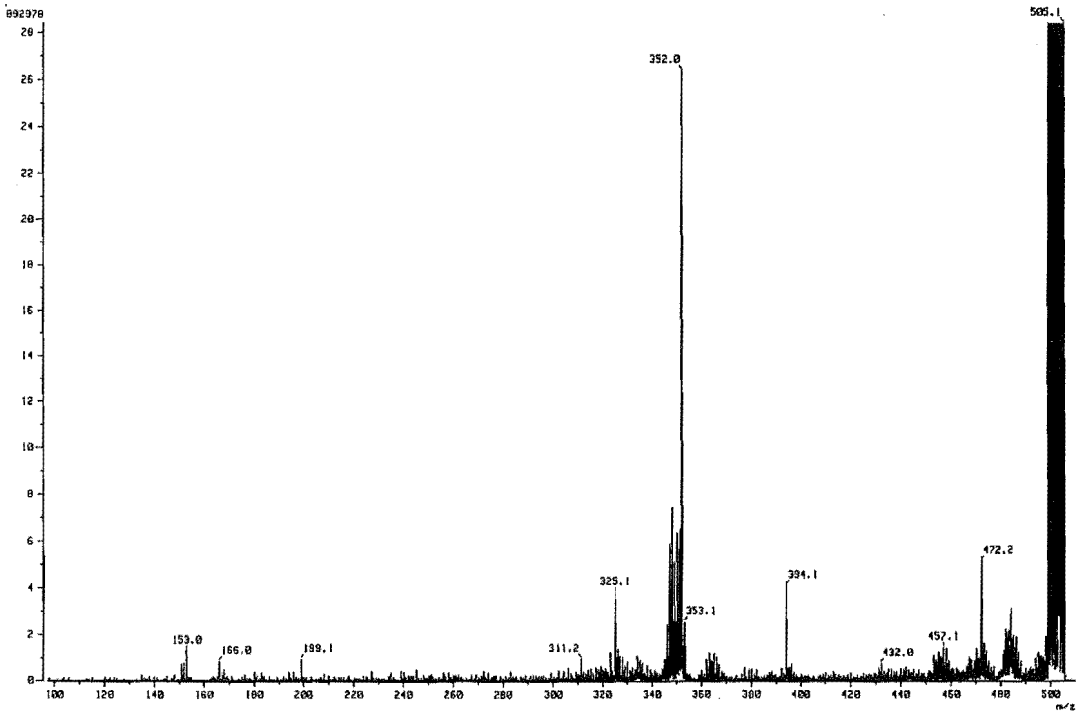
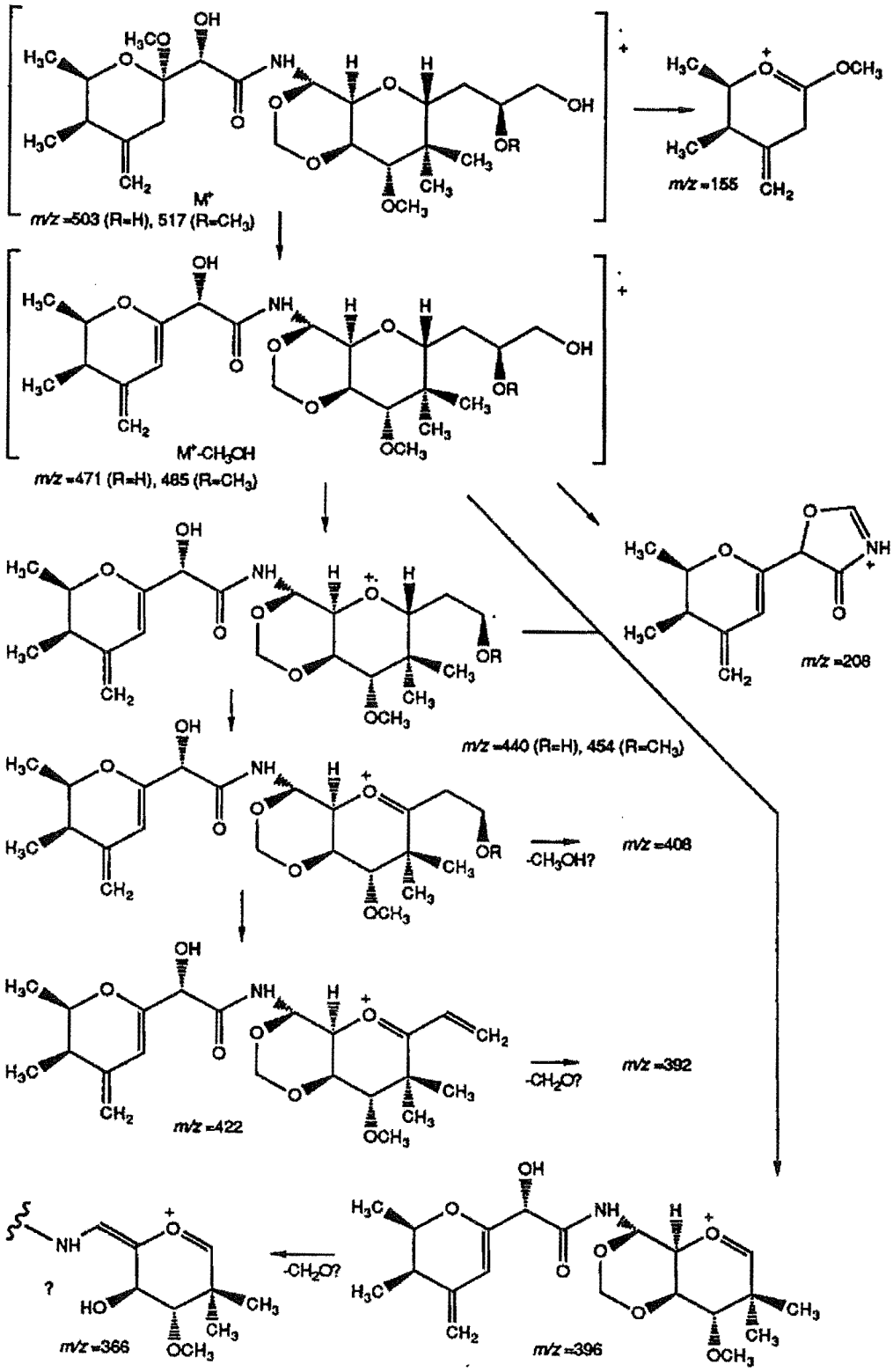


Fig. 3.3.2 Link Scan at *m/z* 502 for Fraction **III.2** from *C. irregularis*



Scheme 3.3.2 EI Mass Spectroscopy Fragmentation of Mycalamide A **(43)**

such as *m/z* 471, 440, 421, 408, 396 (see Scheme 3.3.2) as well as 215, 197, 181, 195 and 87, were recognised as the common peaks in their EIMS spectra (refer to Scheme 3.3.2).

These observations provided supporting evidence for the assignment of identity. The assignment of the ^{1}H NMR spectrum (Fig. 3.3.3) was based on a comparison with that of authentic mycalamide A. Both the chemical shifts and the coupling constants of compound III.2 were consistent with those observed in the ^{1}H-NMR spectrum of mycalamide A.[135] Identical retention times for the mycalamide from the *Callyspongia* sample and authentic mycalamide on RPHPLC also supported the assignment.

3.3.3 Isolation of Cytotoxic Components

In addition to the antiviral components, the chromatography on the flash C-18 column yielded two cytotoxic portions, I.4 and I.6 (see Fig. 3.2.1). Partitioning of fraction I.6 with the HSCCC method concentrated the cytotoxicity into a few fractions, which were combined to give fraction V.1. Further purification of V.1 by LH-20 chromatography produced two active fractions. But, the specific cytotoxicity (mass (mg)/IC_{50} (ng/mL) x 10^4) had decreased from 714 units (V.1) to 23 units (VI.4+VI.5). After passing through a Diol column, the specific activity dropped to just 5 units (see Scheme 3.3.1). This was a significant loss of bioactivity in the course of the column chromatography. Because no acid was used in the above procedures, decomposition, or irreversible absorption of the active components on the stationary phase must have been responsible for the loss.

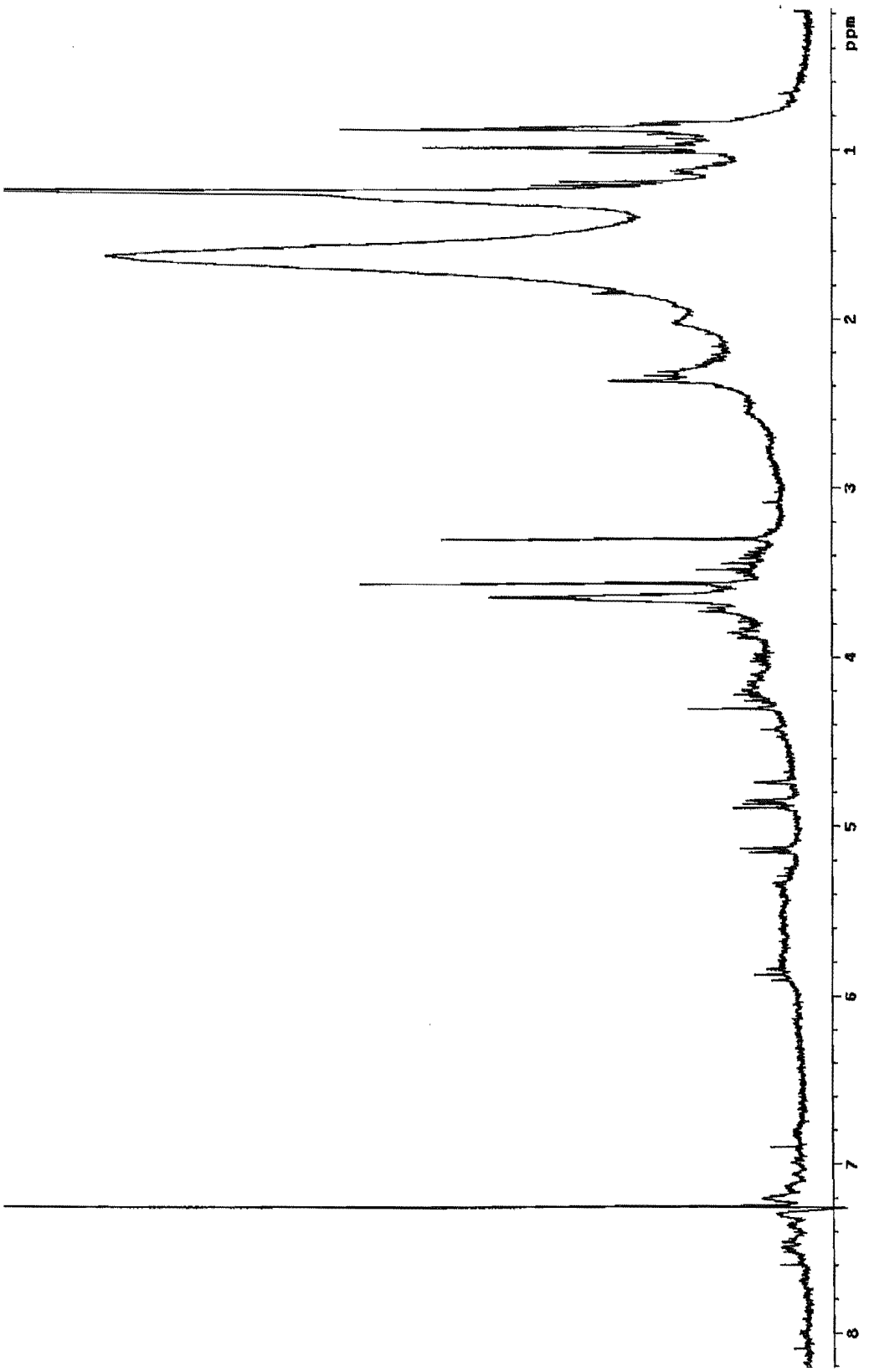


Fig. 3.3.3 ^{1}H NMR (300 MHz) spectrum of fraction **III.2** from *C. irregularis*

The other cytotoxic fraction (I.4) was also explored for location of possible antitumour compounds. Again the HSCCC method gave a good concentration of the bioactivity. However, a total loss of cytotoxicity resulted when a further purification was attempted by RPHPLC. This fraction did not contain mycalamide-like compounds as there was no antiviral activity in I.4. The observed cytotoxicity must be due to another class of compound, that was sensitive either to acid or to the RPHPLC conditions.

These attempted isolations confirmed that HSCCC is a good separation means, especially for dealing with trace amounts of bioactive components which are also unstable. Besides the antiviral compound, mycalamide A, the existence of some other class of cytotoxic components was implied. But, the latter group of compounds are sensitive to the regular isolation conditions. Any future studies on the *Callyspongia* sp. should bear these observations in mind.

3.4 Extraction and Separation of Compounds from *Callyspongia* sp.2

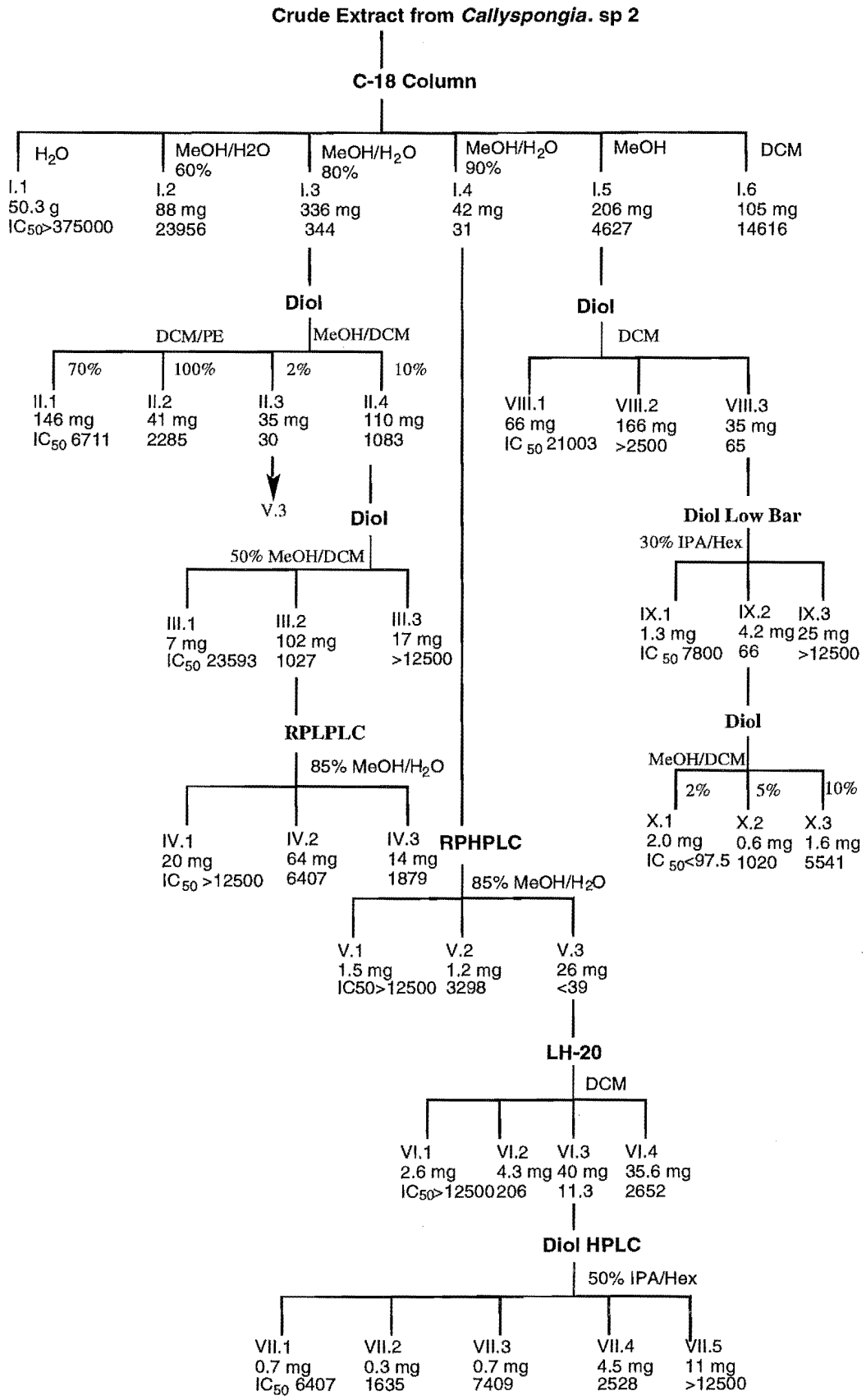
The sponge was collected on the Chatham Rise at a depth of 40-150 m by trawling. The shape and colour of the sponge are similar to *C. irregularis*. However, the initial bioassay exhibited only cytotoxic, but not anti-viral activity. A modified extraction and separation strategy was applied in dealing with this sample.

The frozen sponge was blended and extracted with MeOH, MeOH/CH_2Cl_2 and CH_2Cl_2 in turn. The combined organic extract was dissolved in MeOH/H_2O and partitioned against hexane. After removal of the methanol, the aqueous portion was partitioned against ethyl acetate. These three fractions were then each tested for cytotoxicity against the P388 cell line. The bioassay results (IC_{50} was 431 ng/mL) indicated that the activity was concentrated into the ethyl acetate extract, and no significant bioactivity was present in the aqueous or the hexane fractions. The ethyl acetate extract was chromatographed on a C18 column. The active fractions were combined for further study (see Fig. 3.4.1). The most active fraction displayed an IC_{50} of just 31 ng/mL. But, an HPLC analysis indicated that it was still a complex mixture. It was also found that the cytotoxic components were very "sticky" on the C-18 column, because not much activity was observed in any fractions until the column was stripped with MeOH. After passing through an LH-20 column, the IC_{50} of the most active fraction dropped further down to just 11 ng/mL. When a further purification of this fraction was attempted on a Diol column, most of the bioactivity was lost. The ^1H NMR spectrum of VI.3 (see Scheme 3.4.1) showed three pairs of aromatic signals at δ 7.52, 7.68 and 8.16, and several peaks between δ 3.2-4.4 as well as multiplets at δ

2.10 and 2.38. The active fraction probably contained structures related to polymeric pyridinium alkaloids.

When another of the initial cytotoxic fractions (I.3) was separated on Diol followed by a C-18 column, a loss of bioactivity was again observed. It appears that decomposition of the active compounds accounted for the loss, because the decrease was found from both normal and reverse phase chromatographic procedures.

However, in complete contrast when the Diol was used for the separation of one of the less active fractions (I.5), it worked well, since the specific cytotoxicity increased from 445 units to 5384. But, subsequent chromatography on a Lobar Diol column again resulted in the loss of cytotoxicity (specific unit = 636). This problem was probably caused by irreversible absorption by some contaminants in the stationary phase because the IC_{50} values were similar for fractions VIII.3 and IX.2 (see Scheme 3.4.1). Further purification of IX.2 yielded a cytotoxic fraction with an $IC_{50}<97.5$ (X.1, in Scheme 3.4.1). In the ^1H NMR spectrum of X.1 (Fig. 3.4.1), a singlet at δ 8.03, a doublet at δ 6.76, a doublet of doublets at δ 4.23 and a triplet at δ 3.77 were readily recognised. In addition, the ^1H NMR spectrum also showed other aromatic signals between δ 7.06 and 7.63, while the resonances between δ 3.40 and 4.20 indicated the existence of hydroxyl, ether or amino groups. A broad triplet at δ 5.19 can be attributed to a vinyl proton. From these data, an aromatic moiety with polar functional groups (O or N) was a possibility. The NMR spectrum also implied that this fraction was still a complex mixture. Due to limited amount of material available, this work was stopped.



Scheme 3.4.1 Bioassay-guided separation of the extract from *Callyspongia.* sp.2

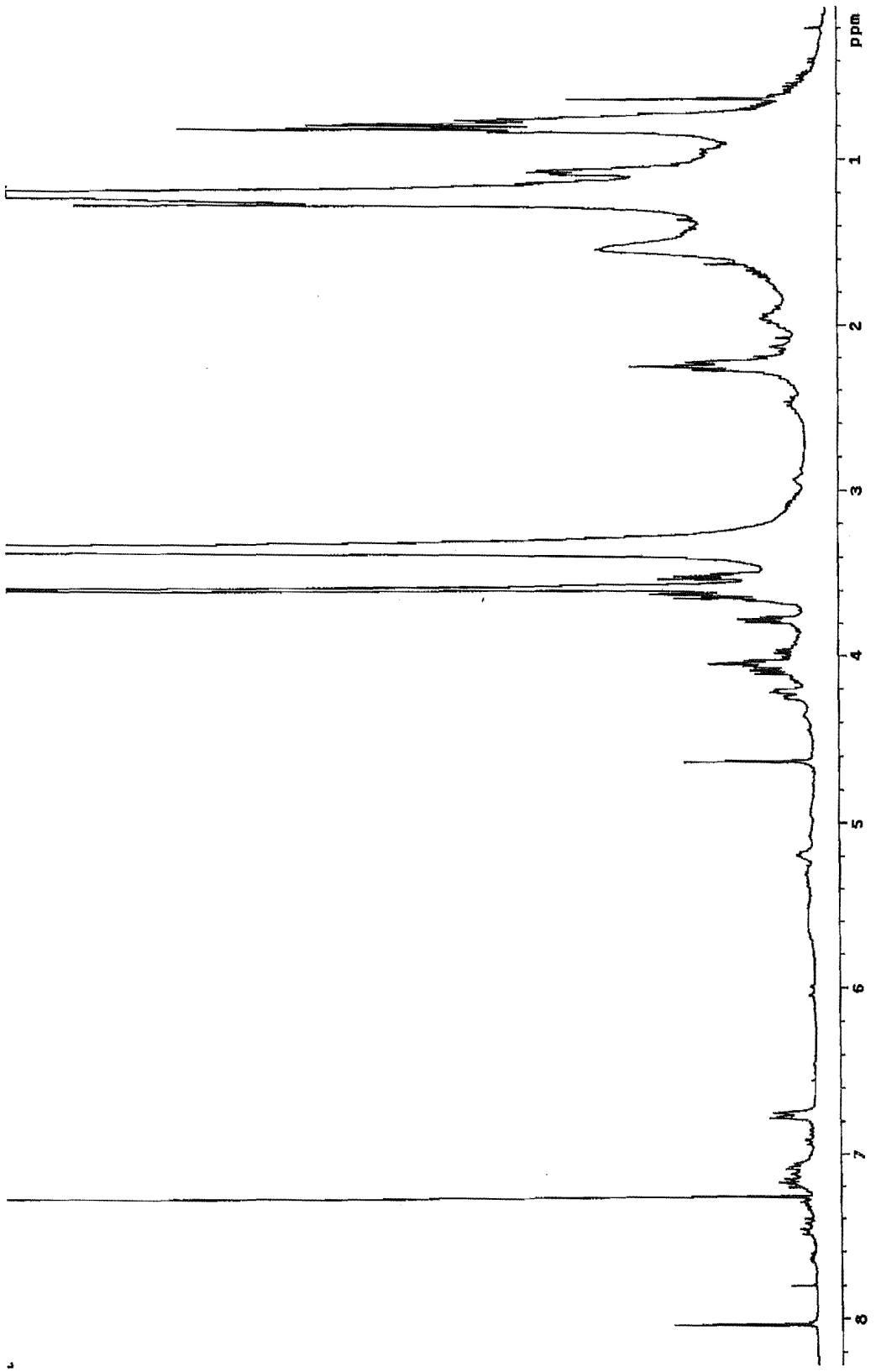


Fig. 3.4.1 ^{1}H NMR (300 MHz) spectrum of fraction X.1 from *Callyspongia*. sp. 2

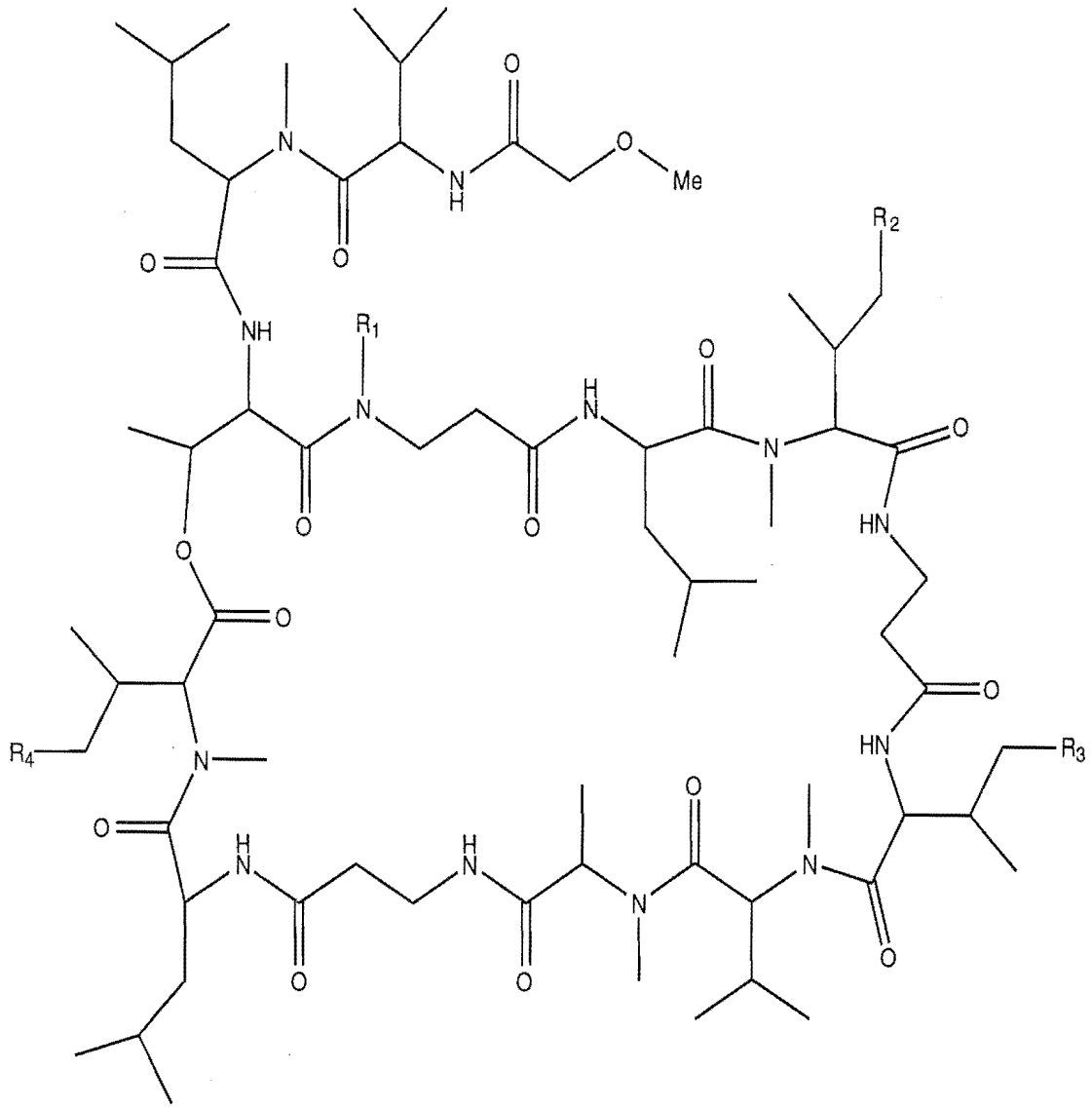
Chapter Four

Theonellapeptolide IIIe

4.1 Introduction

In studies on marine natural products it is rare to encounter several classes of compounds in one species with both interesting structures and biological activities. *Theonella swinhoei* (Theonellidae) and *Lamellomorpha strongylata* are two examples of this rare type of marine sponge. They both contain the cytotoxic compounds the swinholides and the theonellapeptolides, although these two organisms are classified into different sponge orders.

Since the first paper in 1986, Kitagawa's group in Japan has isolated swinholide A and theonellapeptolides Ia, Ib, Ic[136], Id[137], Ie[138] and IId[139] (**44-49**) from the Okinawan marine sponge *Theonella swinhoei*. Swinholide A is a potent cytotoxic dimer macrolide. Compounds **44-49** were found to inhibit development of the fertilised eggs of the sea urchin *Hemicentrotus pulcherrimus* at 2, 2, 2, 50, 10 and 25 μg/mL concentrations respectively.[140] Later, another report indicated that theonellapeptolides Ib, Ic, Id and Ie exhibited cytotoxicity against the L1210 cell line with IC_{50} values of 1.6, 1.3, 2.4 and 1.4 μg/mL.[136] In addition, theonellapeptolides Id and Ie showed ion-transporting activity for either Na^+, K^+, Ca^{2+}, or Na^+ and K^+ ions.[137,140,141] These peptolides are comprised of 13 amino acids. A high ratio of N-methyl amino acids and D-amino acids in the structures are characteristic of this group of compounds. Furthermore, the C-terminus in each of the theonellapeptolides forms an ester bond with the hydroxyl group of threonine, which results in a 37 membered-ring. Because the N-terminus of the peptolides is blocked by a methoxyacetyl group, and also because of the presence of β-alanine, the Edman method was not suitable for sequencing these peptolides. Of necessity, the sequences of the peptolides



44 R_1=H, R_2=CH_3, R_3=H, R_4=CH_3
45 R_1=H, R_2=H, R_3=CH_3, R_4=CH_3
46 R_1=H, R_2=CH_3, R_3=CH_3, R_4=H
47 R_1=H, R_2=CH_3, R_3=CH_3, R_4=CH_3
48 R_1=CH_3, R_2=CH_3, R_3=CH_3, R_4=CH_3
49 R_1=H, R_2=CH_3, R_3=CH_3, R_4=CH_3

Fig. 4.1.1 Structures of theonellapeptolides Ia-Ie and IId

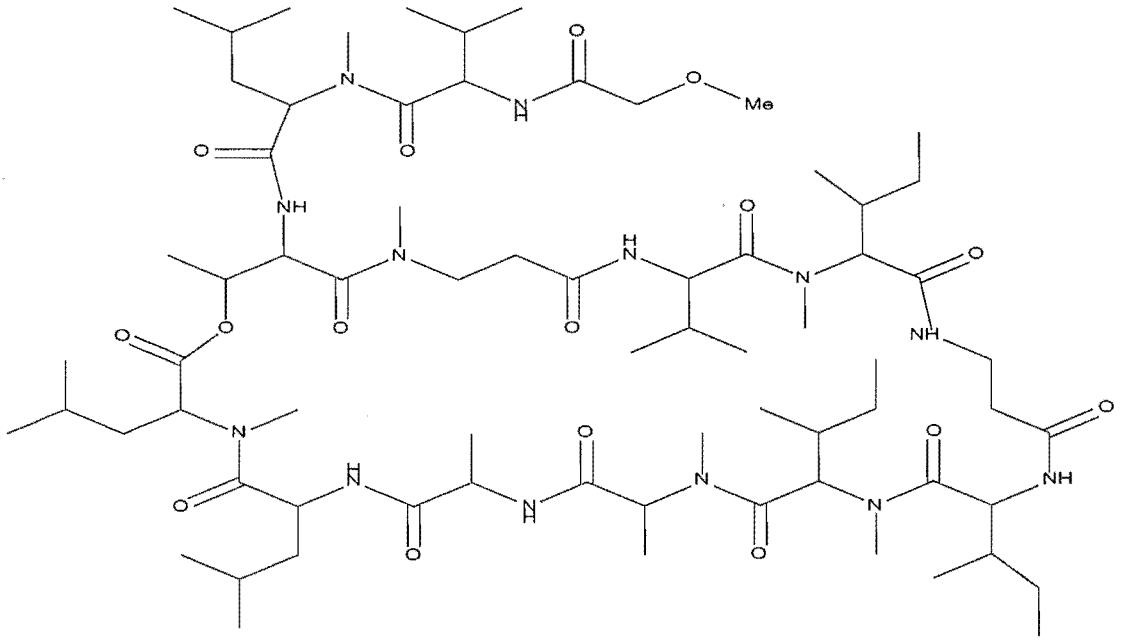
were determined by multi-stepped partial hydrolyses. The absolute configurations of the chiral amino acids were determined *via* reverse phase HPLC chromatography, coupled with CD analysis.

Previous research in bioactive marine natural products undertaken by the Marine Chemistry Group at the University of Canterbury had revealed that deep-water marine organisms are a very promising source of bioactive compounds.[147,124,42] A well-known example is *Lissodendoryx* sp., which was collected using benthic dredging at a depth of 100m. This sponge yielded a series of very potent antineoplastic compounds, the halichondrins. To further explore the potential pharmaceutical resources of deep-water marine organisms, an expedition to the Chatham Rise, 200 km off the East Coast of the South Island, was carried out in early 1995. Bioactivity screening of small scale (2g) extractions on 85 specimens yielded an incidence of cytotoxic and antimicrobial activity almost twice as high (37%) than average of the previous collections. These preliminary assays results confirmed that deep water organisms are a rich source of bioactive compounds.[137] A sponge, *Lamellomorpha strongylata*, showing very potent cytotoxicity in the extracts was chosen for further chemical study.

Dr Eric Dumdei carried out a bioassay-guided separation of *Lamellomorpha strongylata* from the Chatham Rise collection and isolated three distinct classes of cytotoxic compounds. The characterisation of calyculins A, B, E, and F, calyculinamides A and B, and swinholide H, which comprise two of the classes of cytotoxins from this sponge, have been previously reported.[118] This project focused on the purification and structure elucidation of the third class of cytotoxic compounds, the theonellapeptolides.

The frozen sponge was homogenised, extracted with MeOH:DCM (1:1) and subjected to chromatography on a variety of phases. A fraction from LH-20 chromatography showed weak P388 activity (IC_{50}=25 μg/mL). Purification, by repeated RPHPLC, yielded five new active compounds, theonellapeptolides IIIa, IIIb, IIIc, IIId and IIIe, in order of elution (see Section 4.2).

This chapter deals with the structural elucidation of the major peptolide, theonellapeptolide IIIe (**50**). The remaining peptolides will be discussed in Chapter 5 and 6.



Theonellapeptolide IIIe (**50**)

A new structural elucidation strategy, different from that reported previously, was adopted. It was chosen because it proved to be more efficient, less time consuming and more reliable.

The amino acid content of the peptolide was established by GC/MS, following the complete acid hydrolysis of the peptolide and derivatization of the hydrolysate. Next, the cyclic peptolide was methanolysed to yield a ring-opened peptide as the methyl ester, which was sequenced using FAB MS/MS analysis. Concurrently, a series of 1D and 2D NMR spectra were obtained. The NMR techniques, especially 2D TOCSY and HMBC, helped to distinguish between the isobaric amino acids, and constructed the sequence of the peptide. The ROESY spectrum offered more connectivity information, which was used to confirm the entire structure. Finally, the absolute configuration of the constituent amino acids were defined by X-ray crystallography, coupled with a chiral reverse phase HPLC method.

Theonellapeptolide IIIe (**50**), exhibited an MH^+ ion at m/z 1419 in the LRFAB mass spectrum. The molecular formula $C_{71}H_{127}N_{13}O_{16}$ was determined by HRFABMS (m/z 1550.8577, $[M+Cs]^+$, Δ -0.1). The peptidic nature of the fraction was indicated by both the ^1H NMR (doublets between δ 7.18-9.0) and the ^{13}C NMR (resonances between δ 169-176) spectra. A negative result against ninhydrin suggested that the N-terminus was blocked or part of a cyclic peptide, while the observation of a lactone carbonyl stretch (1734 cm^{-1}) in the IR spectrum, coupled with the unsaturation requirements of the molecular formula, indicated that **50** was most likely a peptolide rather than a peptide. As a result of searching the MarinLit[17] database, the structural similarity of this new compound to the known theonellapeptolides was revealed.[136-140]

4.2 Isolation of Theonellapeptolides from *Lamellomorpha strongylata*

The P388 cytotoxic assay on the extract from this sponge showed an off-scale activity [<97.5 x 10^{-6} dilution] in a preliminary assay. Chemical screening carried out by Dr Dumdei on the extract offered some useful information about the properties of the active components. Results from the Diol and C18 cartridges indicated a moderate polarity, while the PEI and CBA cartridges indicated weak anionic properties, and the LH-20 column indicated a relatively large molecular size.

To obtain sufficient material for chemical studies, a larger scale (1 kg) extraction was carried out using MeOH:DCM (1:1). After removal of the organic solvent, the aqueous fraction was acidified to pH 3 and extracted with EtOAc. The combined organic extracts were chromatographed on a C18 column. The activity was concentrated in the fractions eluting off the column between 80% and 100% MeOH. These active fractions were combined and subsequently loaded onto a Diol column, which was eluted in a step-gradient ranging from petroleum ether to methanol. The activity was found in the three bands that eluted with 50-30% petroleum ether in DCM. These three active fractions were each chromatographed on LH-20 using MeOH:DCM (1:1) as eluant to yield three distinct classes of cytotoxic compounds. Most of the biological activity was attributed to calyculins A, B, E, and F, calyculinamides A and B, and swinholide H, which comprised two of the classes of cytotoxins from this sponge. The structures of these compounds were characterised by Dr Dumdei. The third fraction, with weaker P388 activity (IC_{50}=25 μg/mL), eluted just prior to the calyculinamides A and B, and comprised the peptolide fraction.

The active fractions from the LH-20 column on the crude peptolide mixture were combined. Final purification, by repeated RPHPLC of this combined fraction, yielded six cytotoxic compounds, named theonellapeptolides IIIa, IIIb, IIIc, IIId, IIIe and IIIf (Figure 4.2.1). IIIe was the major component, and comprised about 60% of the total peptide mixture.

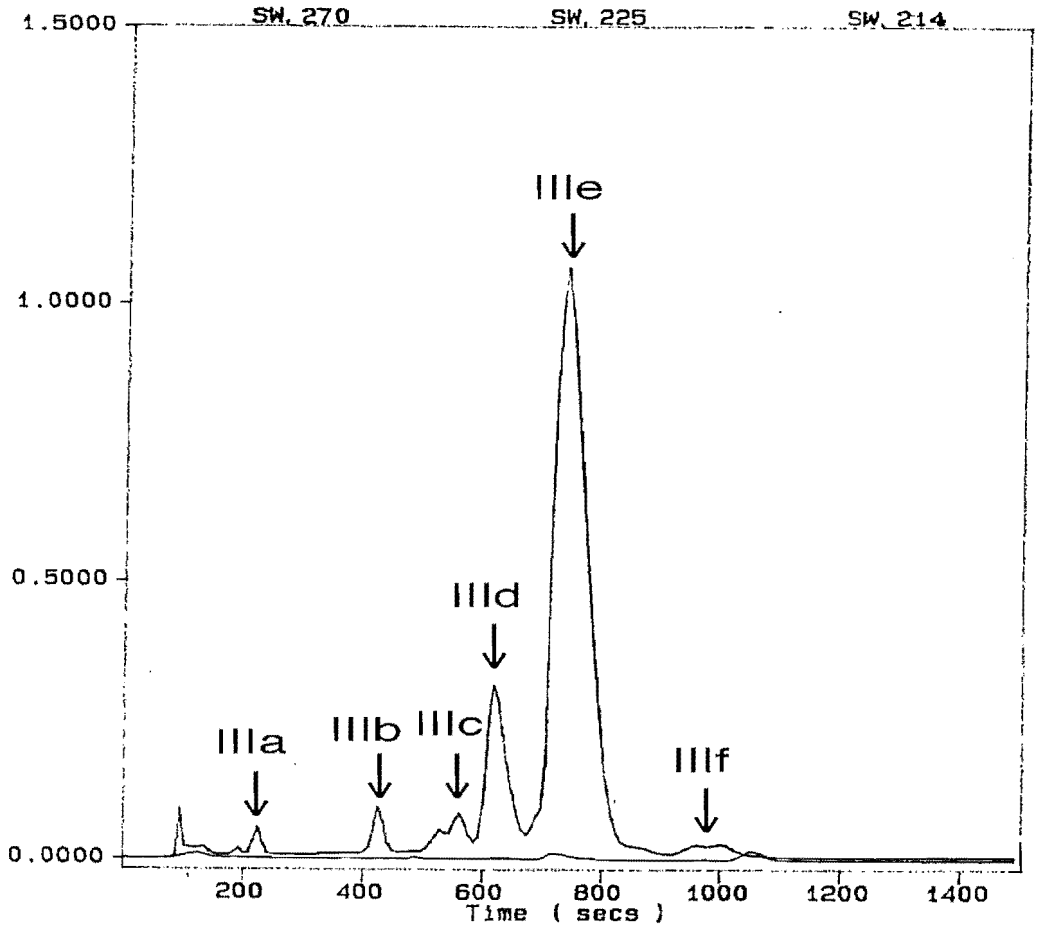


Fig. 4.2.1 The HPLC trace of the peptolide fraction

4.3 Amino Acid Analysis by GC/MS

4.3.1 General Method

Because the ^1H NMR spectrum indicated that theonellapeptolide IIIe contained a high ratio of N-methyl amino acids, these amino acids needed to be identified with more care than the regular amino acids. For the past four decades, various chromatographic techniques, including paper chromatography (PC), thin-layer chromatography (TLC), electrophoresis, high performance liquid chromatography (HPLC), ion-exchange chromatography (IEC) and gas chromatography (GC) have been developed for this purpose. The amino acids were identified in each of these methods through the retention data. The latter three techniques have been the most widely used. However, chromatographic methods alone are not suitable for identification of unusual or unknown amino acids in biological and extraterrestrial materials. This is because there is an absence of structural information. The combination of gas chromatography-mass spectrometry (GC/MS) takes advantage of both the high separation efficiency of GC and the rich structural information from MS, and hence provides assurance on identification of any unexpected or novel amino acids.

The free amino acids themselves are not volatile enough for GC/MS analysis. To increase the volatility of the amino acids, either one or both of the functional groups of the amino acids needs to be derivatised. Previous work has explored a wide range of derivatization methods. Typically, the amino groups can be acylated by C_1-C_5 perfluorinated or non-fluorinated acid anhydrides. The carbonyl group can be derivatised as an ester by C_1-C_5 alcohols, with or without a branch.[94] Of the

possible combinations, N-trifluoroacetyl-n-butyl esters have been most widely used for GC/MS amino acid analysis. Based on both volatility and stability, the N-trifluoroacetyl-n-butyl ester (TAB) derivatives were chosen in this project.

The fragmentation patterns for the TAB derivatives of all protein amino acids and two non-protein amino acids have been established.[143,144] The basic fragmentation pathways for these amino acid are depicted in Fig. 4.3.1.

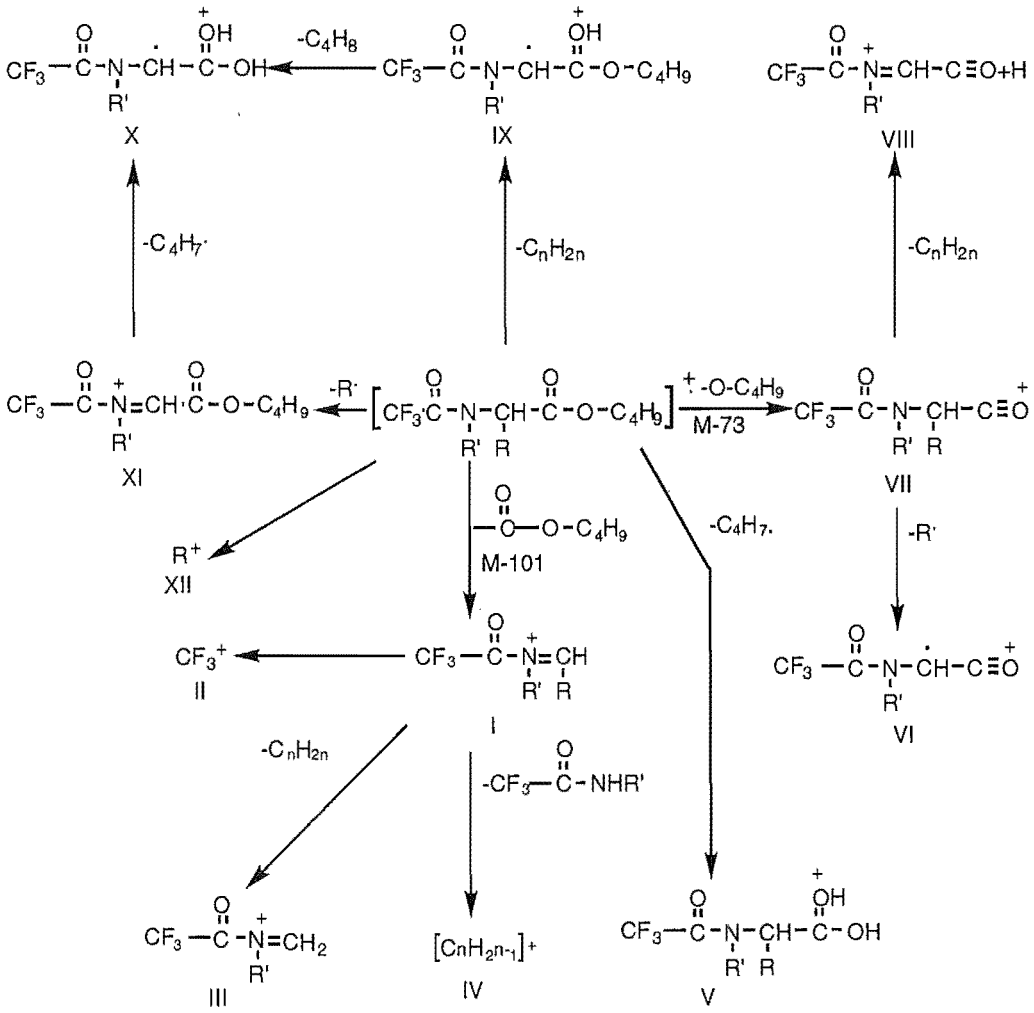


Fig. 4.3.1 Mass spectral fragmentation pathways of TAB α-amino acids

For the aliphatic amino acids, such as glycine, alanine, valine, leucine, isoleucine, norleucine, α-aminobutyric acid, α-aminoisobutyric acid,

sarcosine and related N-methyl variants, cleavage α to the carbonyl group generates either the base peak or the second strongest peak (ion I). Other important fragmentations include cleavage of the bonds between the carbonyl carbon and ester oxygen (ion VII), N and α-C in fragment I (ion IV), α-C and β-C (ion IX and XI), β-C and γ-C (side chain of leucine), and loss of CF_3 (ion II) and C_4H_9 from the side chain.

In this group, leucine, isoleucine and norleucine have the same molecular weight. However, they can be distinguished by different fragmentation pathways of the side chain. For example, the loss of C_3H_7 from the side chain of leucine and C_4H_9 from the side chain of isoleucine respectively are favoured, because secondary radicals are more stable than primary ones. Isoleucine eliminates a butoxy or butene group in conjunction with the cleavage between C2-C3 carbons, yielding higher intensities for the *m/z* 153 and 171 ions than do either leucine or norleucine. Similarly, the cleavage between the C3 and the C4 of leucine is preferred. This generates peaks at *m/z* 140 and 166 in conjunction with the loss of a butoxycarbonyl group or butanol. No major fragments, besides the M-101 (M-$COOC_4H_9$), were found for norleucine.

For hydroxy amino acids, both the amino and hydroxy groups were converted to trifluoroacetyl. The peaks at *m/z* 139 and 153 were observed as the major peaks in the mass spectra of serine and threonine respectively[143]. Also reported were intense *m/z* 138 and 152 peaks. These ions correspond to the loss of the hydroxycarbonyl group plus trifluoroacetic acid. For serine, a M-72 peak rather than a M-73 peak[145] was observed. This ion arises from a double hydrogen rearrangement from the ester portion of the molecule, resulting in elimination of a neutral fragment with the elemental composition of C_4H_8O.[143]

The important fragmentation pathways for β-amino acids are summarised in Fig. 4.3.2.

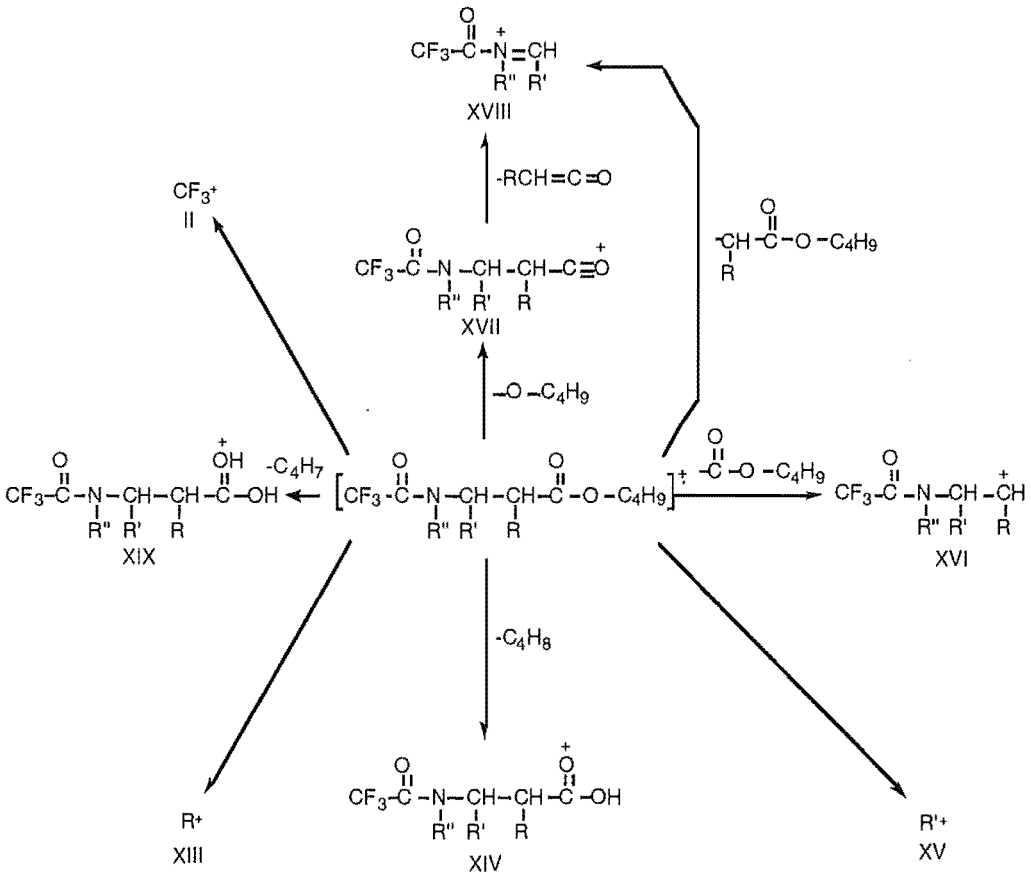
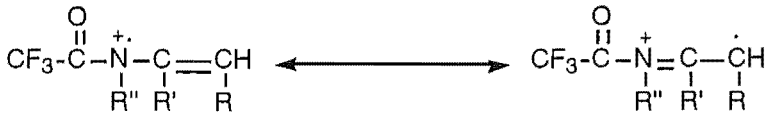


Fig. 4.3.2 Mass spectral fragmentation pathways of TAB β-amino acids

β-amino acids usually give more intense molecular ions than α-amino acids. In contrast to α-amino acids, the fragmentation pathway yielding M-101 (M-$COOC_4H_9$, ion I) is not significant in β-amino acids due to the additional methylene group between the carbonyl and the amide nitrogen. In fact, β-amino acids give a relatively strong M-102 signal resulting from the loss of butoxycarbonyl and a hydrogen. The stability of this ion can be explained as:



Comparing with α-amino acids, β-amino acids also show a more intense M-73 (M-OC_4H_9, ion VII). The intensity observed with β-amino acids can be attributed to the contribution of the following resonance structures:

For all N-methylated α- and β-amino acids, the peak at *m/z* 110 was considered a diagnostic signal due to its absence from the spectra of all non-N-methylated amino acids. This peak has an elemental composition $C_3H_3NF_3$, but its origin is still unknown.

4.3.2 Amino Acid Analysis of Theonellapeptolide IIIe

Theonellapeptolide IIIe was submitted to complete acid hydrolysis. The hydrolysate was converted to N-trifluoroacetyl n-butyl esters using a similar method to that reported.[96] After a preliminary analysis by capillary GC, the derivatives were analysed by GC/MS using the same type of column. A total of ten main GC peaks were recognised (Fig. 4.4.3). They were labelled in order of increasing retention time.

Peak 1 exhibited a low intensity molecular ion at *m/z* 255. The peak at *m/z* 110 is characteristic of an N-methyl amino acid. Signals at *m/z* 154 and 182 corresponded to the loss of butoxycarbonyl and a butyl group respectively. Thus, peak 1 was identified as N-methylalanine.

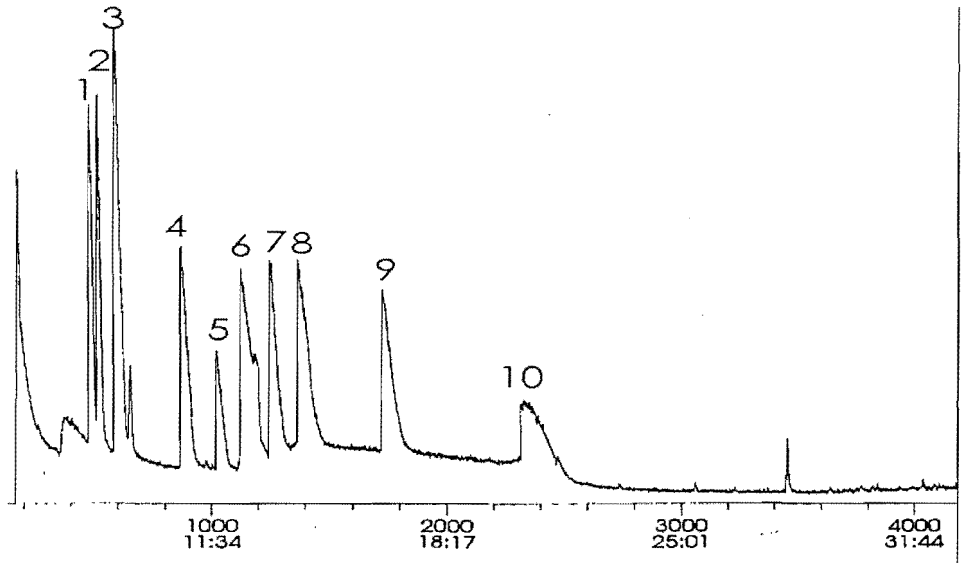


Fig. 4.3.3 GC/MS trace of the N-TFA n-butyl esters of the hydrolysate from theonellapeptolide IIIe

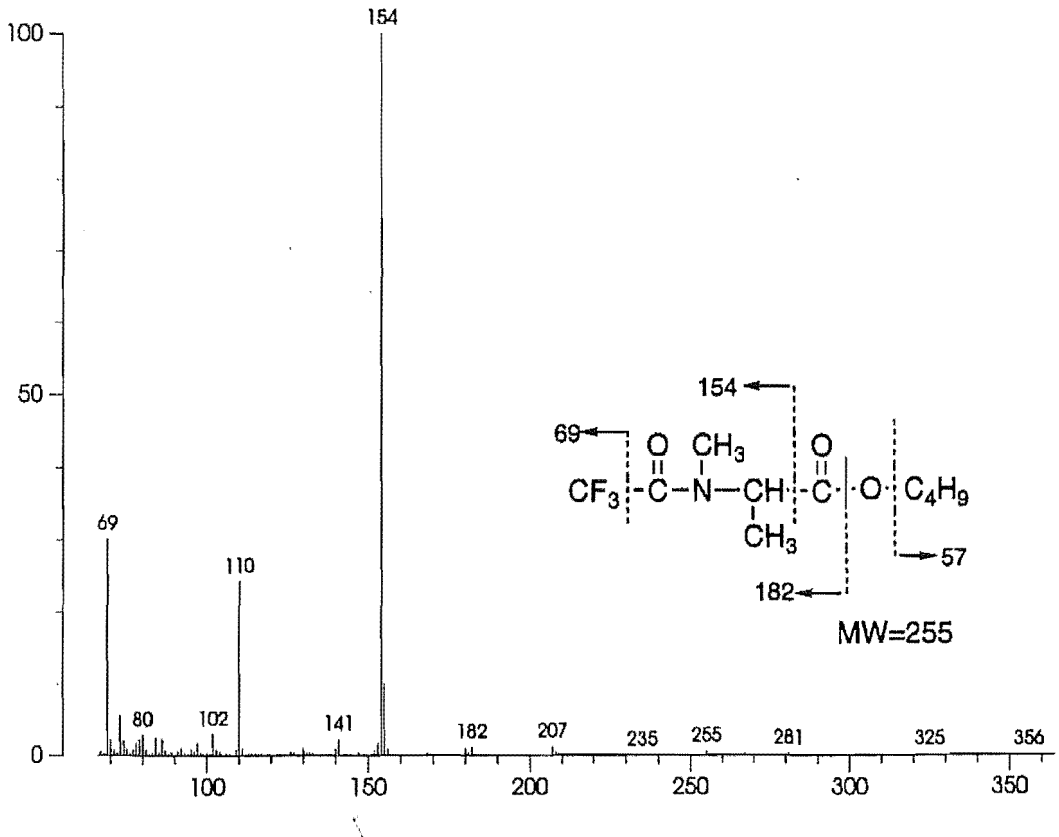


Fig. 4.3.4 Mass spectral fragmentation of N-TFA-n-butyl ester of N-methylalanine

In the mass spectrum of peak 2, the base peak was at *m/z* 140. This is a characteristic peak for alanine, sarcosine and N-methyl-β-alanine. The low intensity ion at *m/z* 110, co-occuring with the *m/z* 154 ion, implied contamination from N-methylalanine. This suggested that peak 2 is alanine, rather than sarcosine and N-methyl-β-alanine. The ions at *m/z* 168 and 186 arise from M-73 (C_4H_9O) and M-55 (C_4H_7). The peak at *m/z* 126 can be attributed to ion III (Fig. 4.3.1).

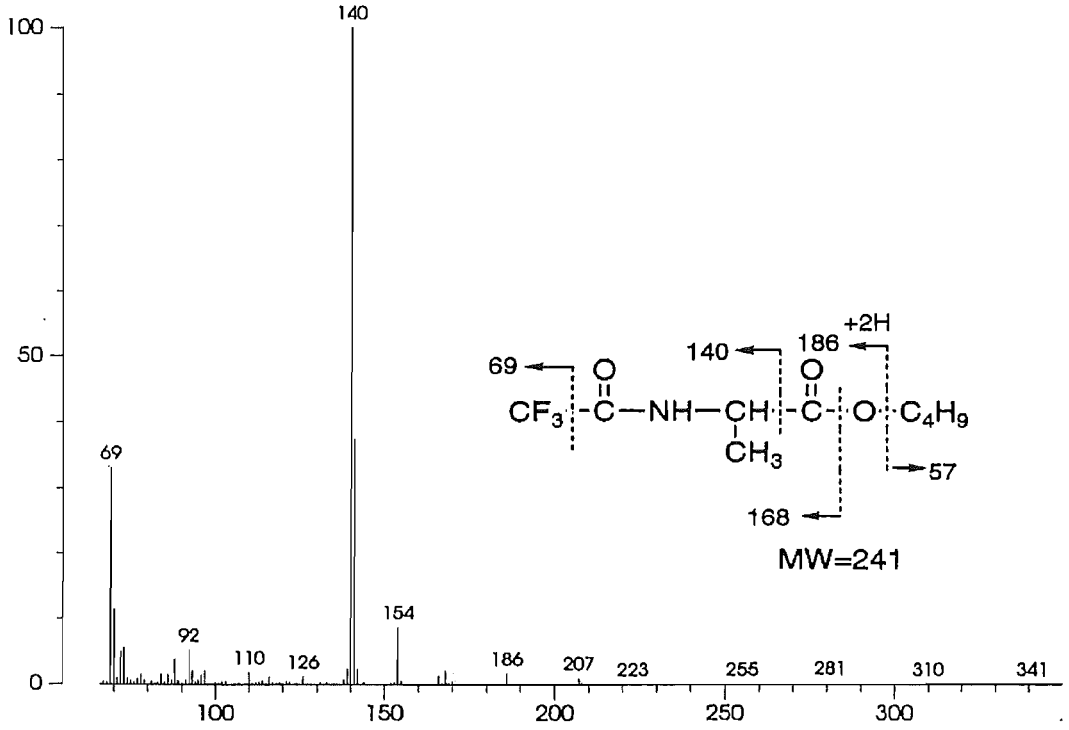


Fig. 4.3.5 Mass spectral fragmentation of N-TFA-n-butyl ester of alanine

Peak 3 was identified as valine by comparison of MS data with that reported by Leimer *et al.*[143] In the spectra of both peak 3 and valine, the base peak at *m/z* 168 (I in Fig. 4.3.1), corresponding to cleavage α to the carbonyl carbon, was quite distinctive. The relatively strong *m/z* 153 ion was attributed to the elimination of an isopropyl group from *m/z* 196, which corresponded to the loss of butoxy (M-73). The ion at *m/z* 227 arises from the loss of C_3H_6 from the side chain of the amino acid. However, the fragmentation mechanisms were not clear.

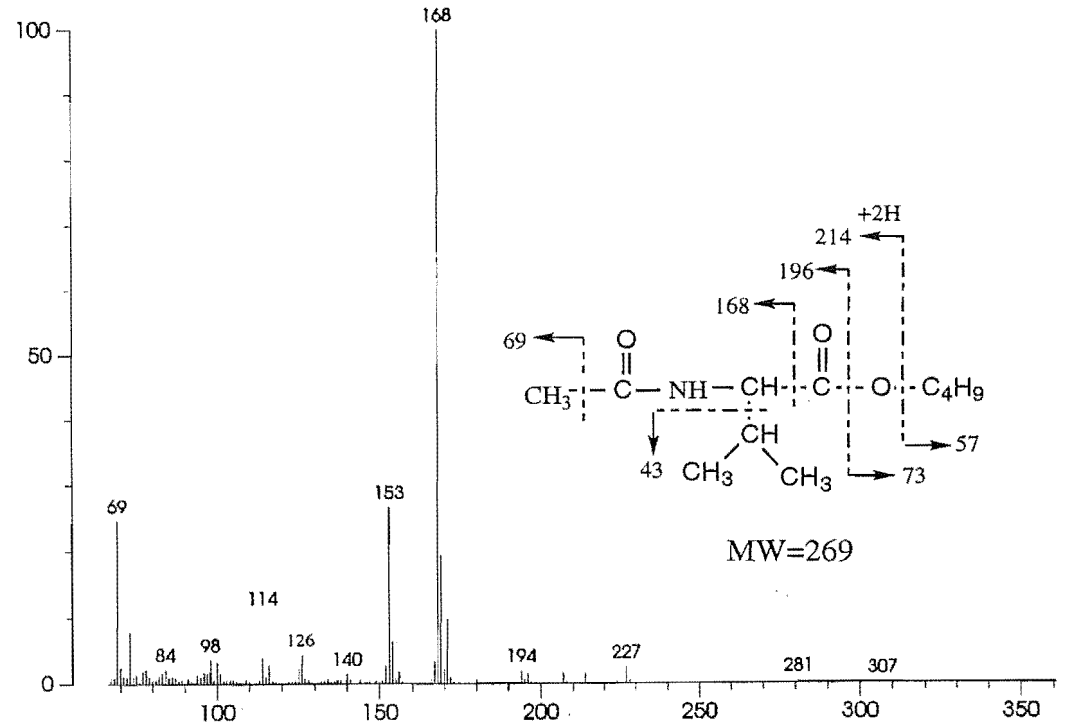


Fig. 4.3.6 Mass spectral fragmentation of N-TFA-n-butyl ester of valine

The major fragmentation patterns for peak 4, shown in Fig. 4.3.7, showed all the characteristics of an α-amino acid. The strong signal at *m/z* 182 corresponded to M-101 (ion I in Fig. 4.3.1), which then eliminated C_4H_8 from the side chain of the amino acid to yield the equivalent of structure III at *m/z* 126. The peak at *m/z* 171 was formed by loss of C_4H_8 from *m/z* 227, (cf. ion V in Fig. 4.3.1). Further examination led to the conclusion that the amino acid in question was isoleucine because of a relatively strong *m/z* 153. This ion corresponds to further elimination of a stable isobutyl group from the ion equivalent to VII.

All the major signals in the spectrum of peak 5 (see Fig. 4.3.8) were 14 mass units more than the corresponding fragments of isoleucine (Fig. 4.3.7).

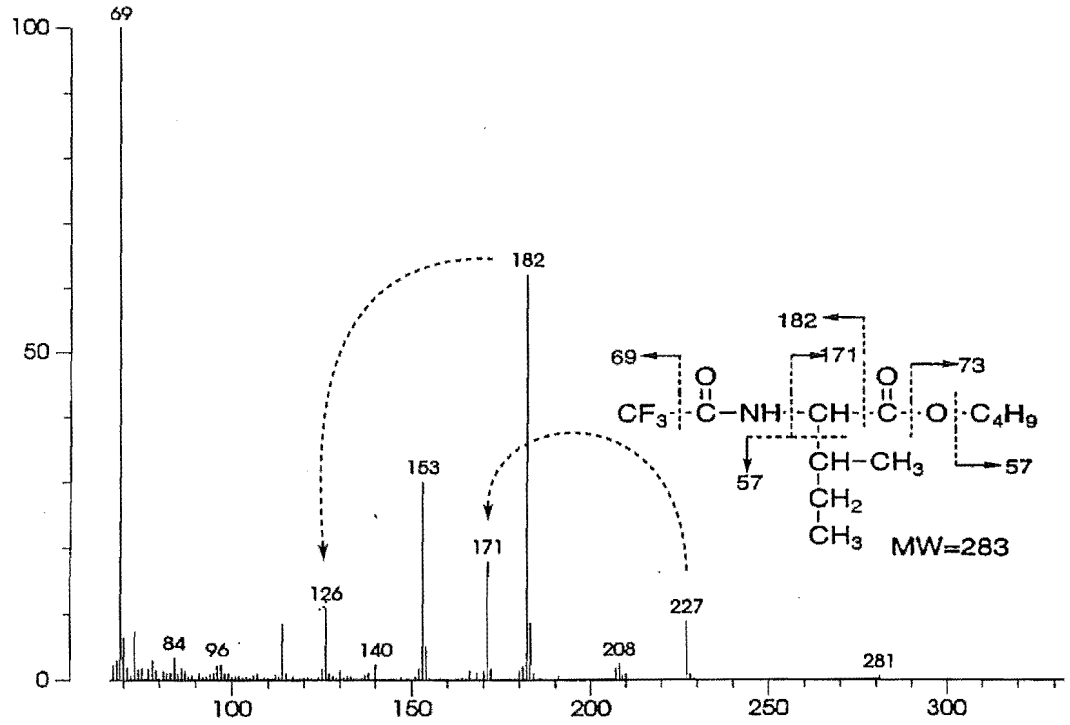


Fig. 4.3.7 Mass spectral fragmentation of N-TFA-n-butyl ester of isoleucine

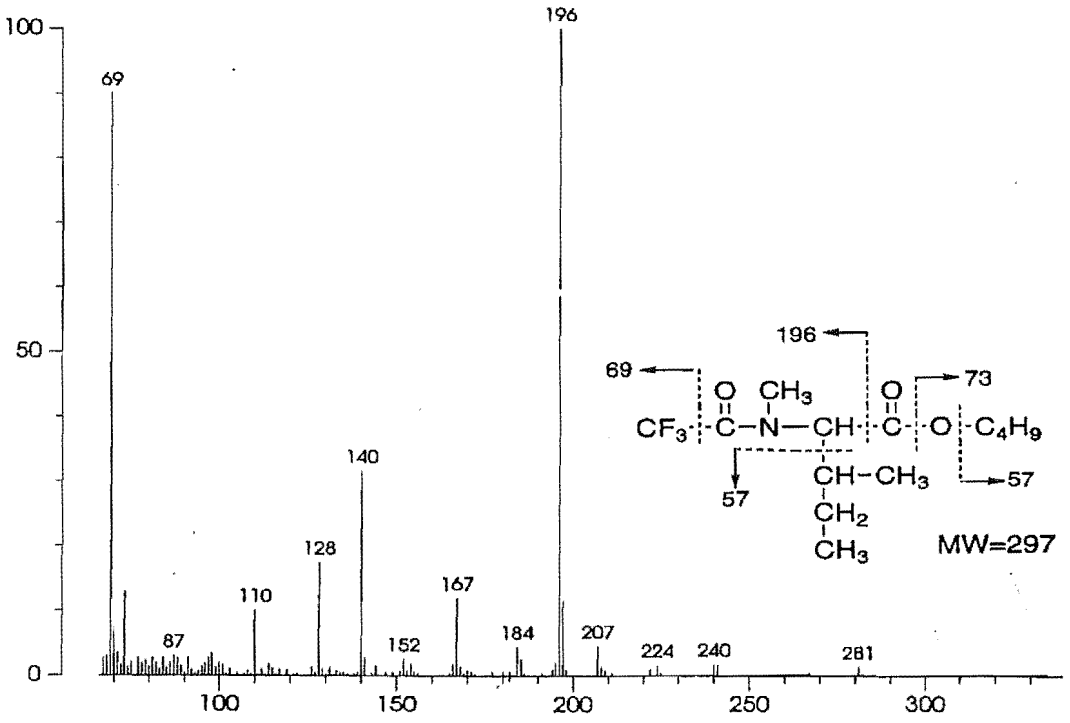


Fig. 4.3.8 Mass spectral fragmentation of N-TFA-n-butyl ester of N-methylisoleucine

However, an *m/z* 110 ion was also observed. These data suggest that peak 5 is N-methylisoleucine. The ion at *m/z* 224, which can be explained by structure (VII), lends support to this assignment.

The mass spectrum of peak 6 (Fig. 4.3.9) displayed a similar fragmentation pattern to that for the N-TFA-n-butyl ester of N-methylisoleucine, suggesting another N-methyl amino acid with the same molecular mass as peak 5. A relatively strong *m/z* 154 *vs* 140 ion implied an N-methylleucine structure. The strong *m/z* 154 ion can be explained by the loss of the isopropyl group from the ion I through cleavage of the bond between the C3 and C4 carbons, while the relatively low intensity of the *m/z* 140 ion is due to the loss of an unstable primary butyl group from ion I by cleavage of the bond between carbons 2 and 3 (see Fig. 4.3.1). In this way, peak 6 was assigned to N-methylleucine.

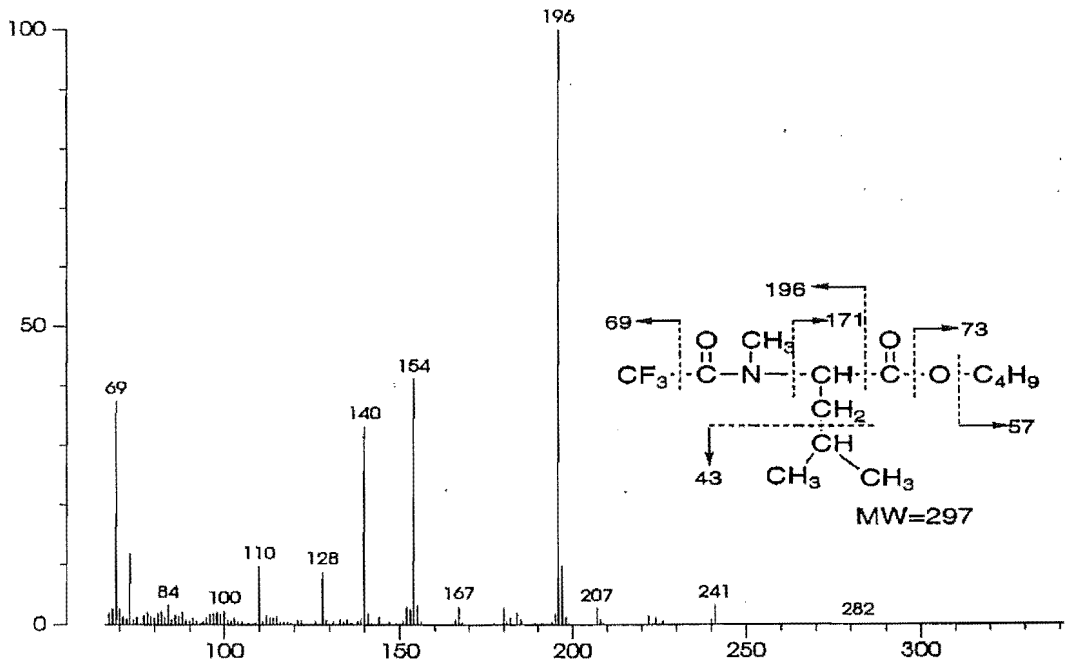


Fig. 4.3.9 Mass spectral fragmentation of N-TFA-n-butyl ester of N-methylleucine

The spectrum of peak 7 shows a molecular ion at *m/z* 255. The molecular mass was the same as that of N-methylalanine, but the fragmentation patterns are different. The base peak was observed at *m/z* 154 in the former, but was at *m/z* 140 in the latter, a 14 mass unit difference. Apparently the most favoured cleavage was shifted by a CH_2 unit. This was consistent with the characteristic fragmentation pathway of β-amino acids. A strong M-102 at *m/z* 153 rather than 154, and the observation of an ion at *m/z* 110 confirmed this thinking. This amino acid was assigned as N-methyl-β-alanine.

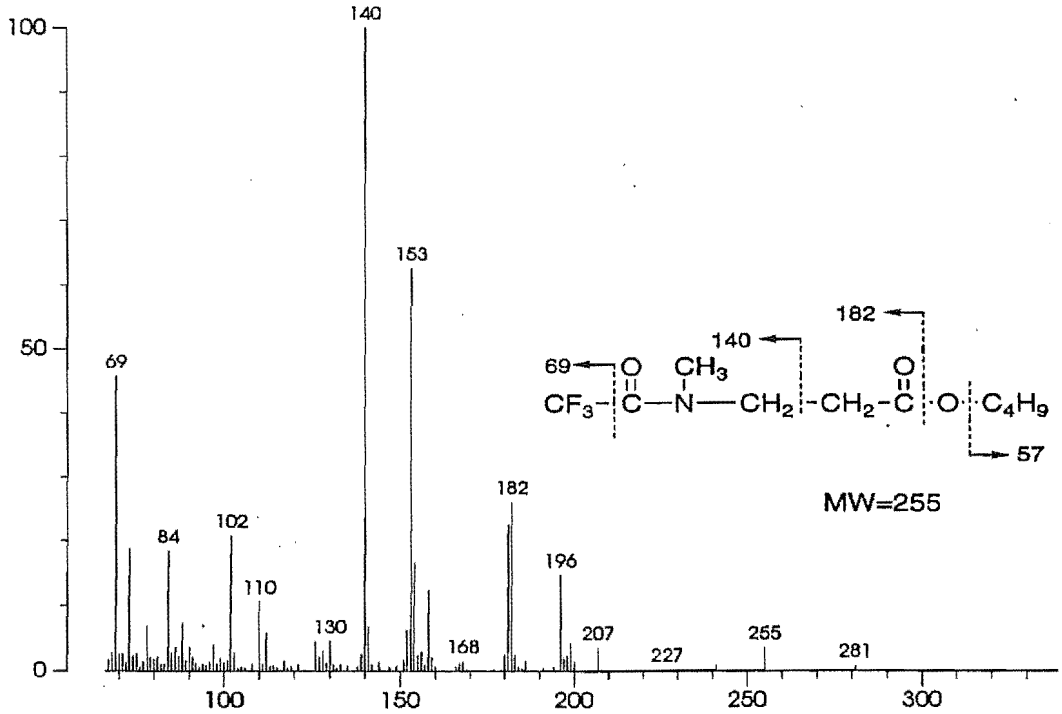


Fig. 4.3.10 Mass spectral fragmentation of N-TFA-n-butyl ester of N-methyl-β-alanine

Peak 8 was identified as the N-TFA-n-butyl ester of leucine based on the following analysis of the dominant fragmentation ions (see Fig. 4.3.11). The second strongest peak at *m/z* 182 in the spectrum was formed by cleavage α to the carbonyl carbon. The peak at M-101, due to the loss of C_4H_8, yielded the *m/z* 140 ion. Successive loss of C_4H_9, C_3H_7 and CF_3 from the molecular ion resulted in *m/z* 114. Ions at *m/z* 126, 153, 171 and 227 can be explained by the structures III, VI, X and IX respectively (Fig. 4.3.1).

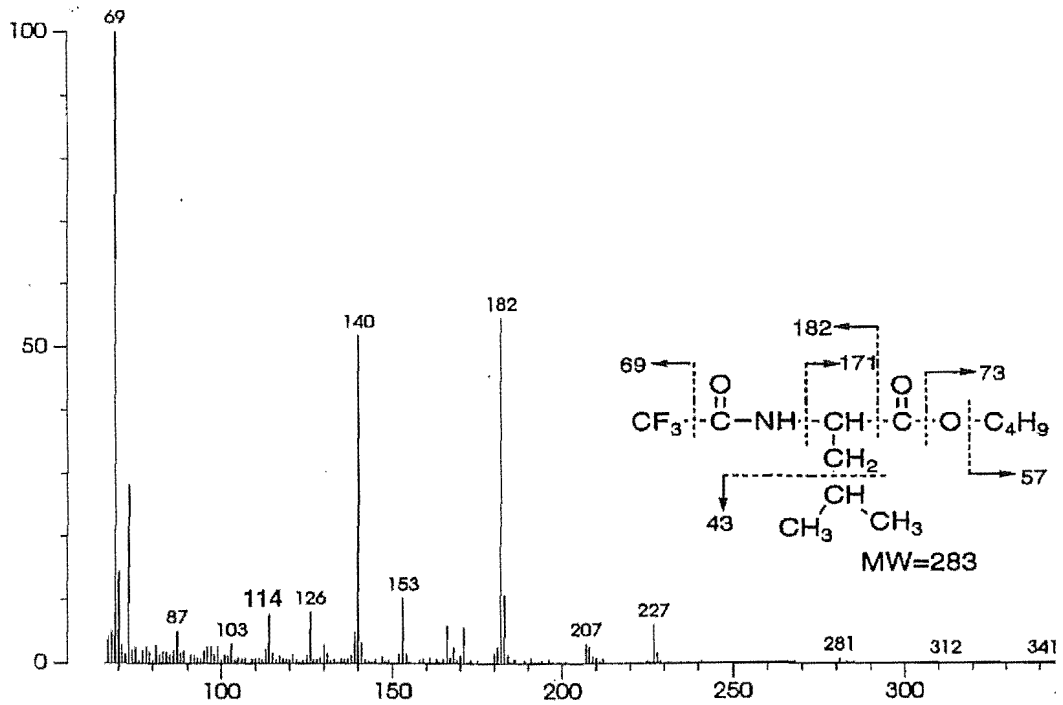


Fig. 4.3.11 Mass spectral fragmentation of N-TFA-n-butyl ester of leucine

A comparison of the mass spectrum with published MS data led to the identification of peak 9. There was a high correlation between published data for β-alanine and peak 9 (Fig. 4.3.12), except for the ion at *m/z* 207. (Because this ion appeared in all spectra of the derivatives of the hydrolysate components, it was possibly a contaminant.) The major fragments, ions at *m/z* 116, 126, 139, 168 and 186, can be reasonably explained as M-CF_3-CO-N=CH_2, β-cleavage of the amide nitrogen, M-102, M-C_4H_9O and M-55.

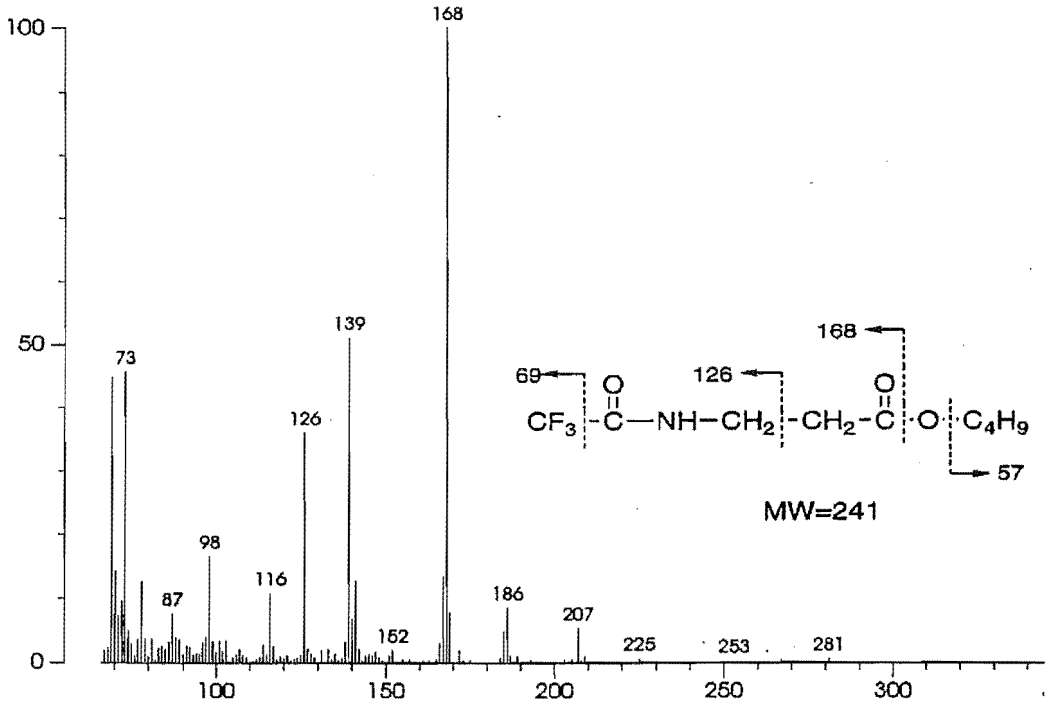


Fig. 4.3.12 Mass spectral fragmentation of N-TFA-n-butyl ester of β-alanine

The GC behaviour of the last amino acid derivative indicates that its polarity was much higher than others in the hydrolysate. Although a molecular ion could not be detected, the ions at *m/z* 281, 266 and 227 can potentially be assigned as arising from M-72, M-101 and M-140. This evidence, coupled with the very strong signal at *m/z* 153 and 152 suggested that this peak corresponded to threonine. The cleavage between C2 and C3 accompanied by a hydrogen transfer to the carbonyl group would yield the ion *m/z* 227. Elimination of a stable secondary CF_3-CO-O-CH-CH_3 radical in conjunction with the loss of C_4H_9O (*m/z* 73) accounts for the intense *m/z* 153. Ions at *m/z* 141, 179 and 198 can be attributed to cleavage β to the carbonyl carbon with the charge retained on the CF_3-CO-O-CH-CH_3 fragment, elimination of trifluoroacetic acid from M^+ followed by the loss of butanol, and C_4H_7 respectively.

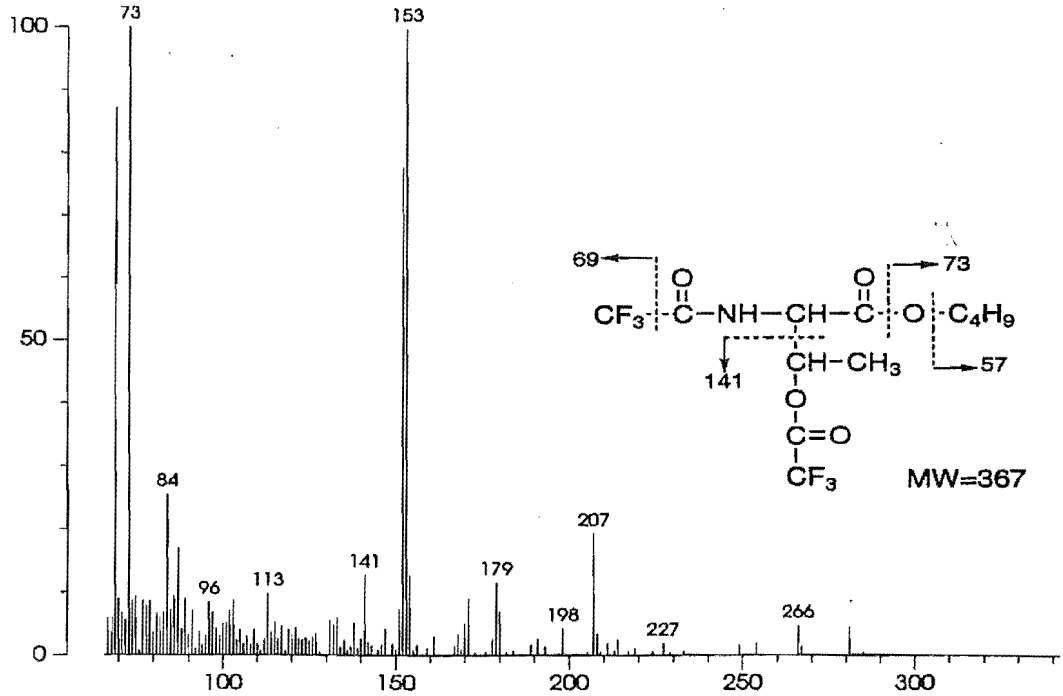


Fig. 4.3.13 Mass spectral fragmentation of N-TFA-n-butyl ester of threonine

Examination of the relative integral areas in the GC trace revealed that peaks 3, 5 and 6 each comprised two moles of amino acids. This result, coupled with the above mass spectral analysis, indicates that theonellapeptolide IIIe was comprised of 13 amino acids, which included alanine, N-methylalanine, β-alanine, N-methyl-β-alanine, isoleucine, N-methylisoleucine (x2), leucine, N-methylleucine (x2), threonine and valine (x2).

4.4 Sequencing the Ring-opened Peptide

The amino-acid sequence plays a vital role in the structural studies of peptides. Sequencing methods, such as the Edman degradation, enzyme or acid partial hydrolysis, GC/MS and MS, have been developed. Of these, the Edman method is the best known and most widely applied. Several automated sequencers based on the Edman sequence are commercially available. However, some limitations of the method itself can result in the failure of sequential Edman degradation. This method can not be used when the N-terminus of the peptide is blocked or modified, or unusual amino acids are present.

Because the theonellapeptolides contain β-alanine moieties and have their N-termini blocked, they can not be sequenced by the Edman degradation. However, Kitagawa's group successfully sequenced six theonellapeptolides using partial hydrolysis in conjunction with MS and HPLC analysis. Their strategy was to separate fragments from partial acid hydrolysis by HPLC, and then to elucidate the structures of the fragments by MS and NMR. The larger fragments were submitted to repeated partial hydrolysis and spectral analysis, until the whole sequence was established. To sequence theonellapeptolide IIIe, the same method was tried. The peptide was hydrolysed with 30% trifluoroacetic acid (TFA) at 110°C for 40 minutes. The hydrolysate was then analysed by both column chromatography and TLC. It transpired that only a very small proportion of the peptolide had been hydrolysed. To find better reaction conditions a series of experiments using different concentrations of HCl with varied reaction temperatures or reaction times were carried out. All hydrolysates were analysed by HPLC. It was found that most of the reactions produced more than ten fragments, but the starting material was

the largest fraction. In order to clarify the situation, the fragments from one such reaction were analysed by FABMS. The spectra did not provide much information on the sequence, as the expected cleavages were not observed. The spectra were obscured by complex fragmentation. Considering both the time component (structural elucidation for one typical peptolide required separation and characterisation of 19 fragments) and the sample cost, a different strategy was undertaken.

Over the past decade, FABMS/MS has emerged as a powerful sequencing technique. This has gained more and more attention from both biologists and chemists. The advantage of FABMS over conventional techniques is that unusual and modified amino acid residues are readily identified. Mixtures containing several peptides are amenable to analysis, and since there is no requirement for a free N-terminus, cyclic peptides and N-terminally blocked materials can also be characterised. The working principle of FABMS can be simply described. The sample molecule is dispersed in a matrix, and bombarded with either fast atoms or ions to cleave the molecular ion (or $[M+H]^+$) and produce daughter ions, which are recorded by a detector. However, a number of steps may occur between those two types of ions. A daughter ion can possibly have more than one origin. The peptide can not be sequenced unambiguously by a single FABMS spectrum, because it offers little structurally significant information. This disadvantage can be remedied by tandem mass spectrometry (MS/MS). The MS/MS technique not only gives clues as to the relationship between precursor (parent) and product (daughter) ions, but also increases fragmentation abundance and eliminates the interferences from matrix fragmentation.

The possible fragmentation pathways of a peptide in FABMS are outlined in Fig. 4.4.1. The nomenclature utilised in this manuscript for the amino acid sequence-determining fragment ions follows that proposed by Roepstorff and Fohlman[146]. The three possible cleavage points of the peptide backbone are labelled A, B and C when the charge is retained on the N-terminus of the peptide, and X, Y and Z when the charge is retained on the C-terminus. The numbering (subscript *i*) indicates the peptide bond cleaved, counting from the N- or C-terminus respectively. The number of hydrogens transferred to or lost from the noted fragment are indicated by a corresponding number of primes placed to the right or left of the letter respectively.

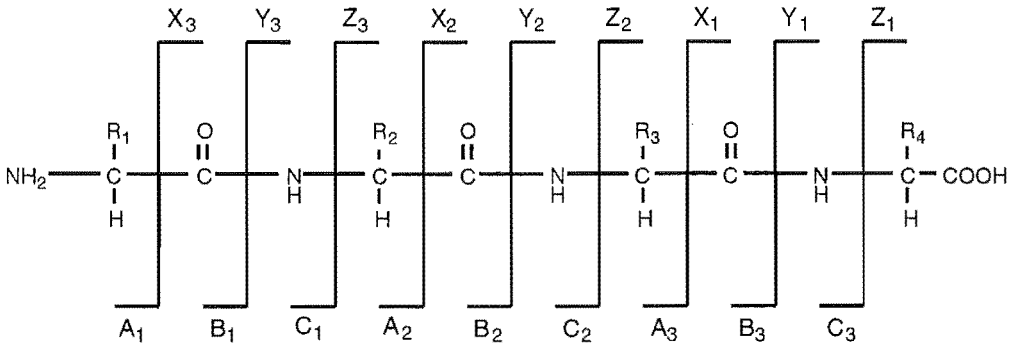


Fig. 4.4.1 Nomenclature used for the amino acid sequence-determining fragment ions

The peptolide IIIe, as such, was not directly suitable for FABMS/MS analysis. To make a suitable form for sequencing, the lactone bond of theonellapeptolide IIIe was hydrolysed with KOH. After purification on reverse phase HPLC, the major product, corresponding to a linear peptide, was submitted to FABMS analysis. Unfortunately, neither the $(M+H)^+$ nor intense signals in the high mass range (mass value>600) could be recognised. There was not much information that could be

extracted. These problems probably resulted from a strong competition for ionisation or inhibition from salts[147,148]. Because the linear peptide was generated from basic hydrolysis, a carboxylate anion probably existed. The anion would suppress the protonation of the molecule, hence causing a very low intensity of $(M+H)^+$ to be observed. This assumption has been supported by observations from other researchers.[149,150,151] In the positive mode of FABMS, basic peptides generally give good signals, whereas the acidic peptides sometimes can not be detected.[152]

To ascertain whether blocking the carboxylate group would improve the quality of the spectrum, the methyl ester of the ring-opened peptide was prepared by methanolysis of the peptolide with $NaOCH_3/CH_3OH$. The hydrolysate was separated by RPHPLC and produced four fractions. ^1H NMR were obtained for all fractions. The major product (**51**) showed peptidic character. An NMR comparison with the starting material revealed that all amide protons and N-methyl protons had obvious shifts, and that a new methyl signal had appeared at δ 3.55, attributable to a methyl ester. HRFABMS gave the molecular formula $C_{72}H_{131}O_{17}N_{13}$. This molecular mass was 32 amu higher than theonellapeptolide IIIe, corresponding to addition of a CH_3OH to the parent. This result further confirmed the formation of the methyl ester.

The low resolution FABMS spectrum (Fig. 4.4.2) of the linear methyl ester provided much more information than that of the KOH hydrolysate. The highest-mass peak in the FAB spectrum, using glycerol as the matrix, was *m/z* 1451. This was assumed to be $(M+H)^+$. In a spectrum using NOBA as the matrix, ions at *m/z* 1603, 1489 and 1473, which corresponded to M+23 (Na^+), M+39 (K^+) and M+153 (NOBA), were also

These ions helped to confirm that the ion at *m/z* 1451 was indeed $(M+H)^+$.

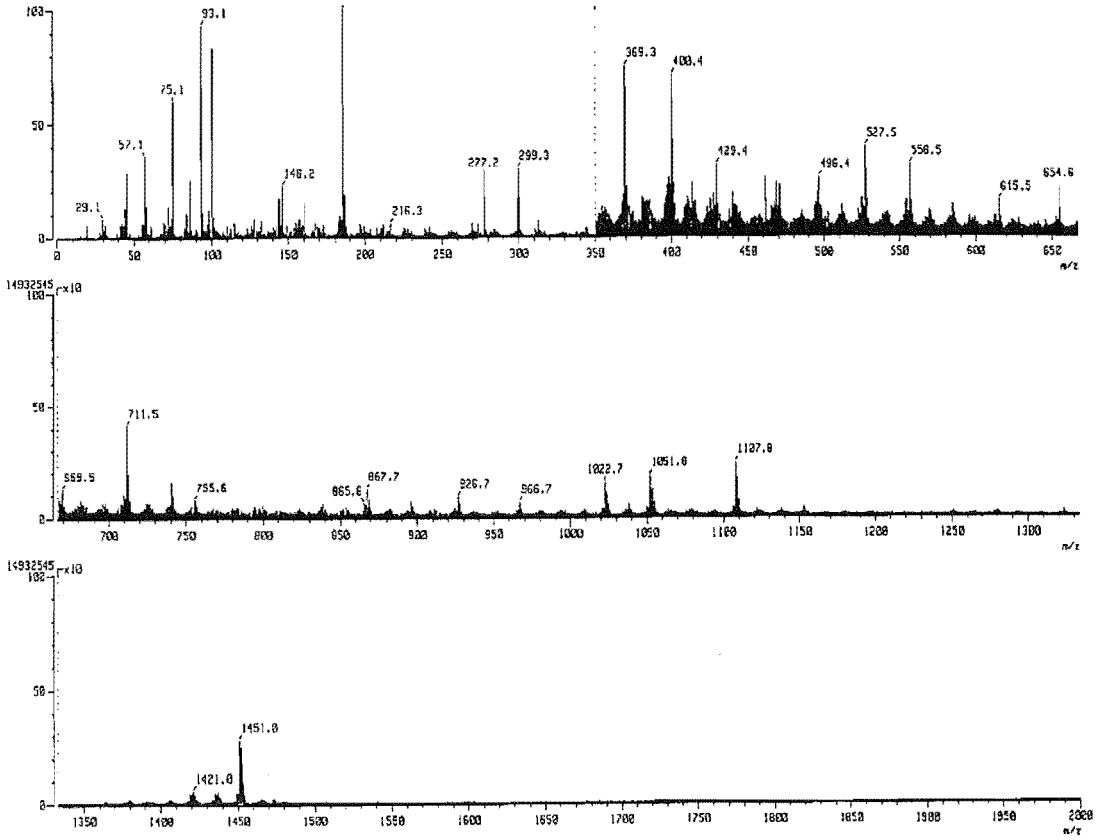
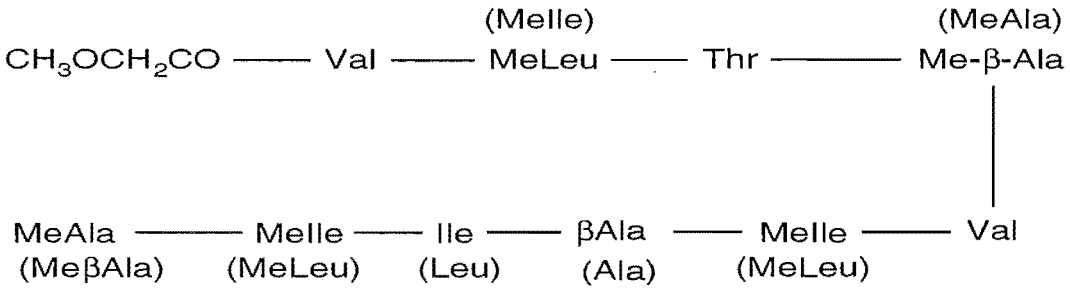


Fig. 4.4.2 FABMS spectrum of compound **51**

The first reasonably strong peak below m/z 1451 (Fig. 4.4.2) appeared at *m/z* 1107, with a mass gap of 343 amu from the $(M+H)^+$. This mass difference can be considered as loss of a fragment comprising three amino acids. From *m/z* 1107, the ions at *m/z* 1022, 895, 711, 584, 400, 299 and 172 can be viewed as the loss of amino acid residues one by one or possibly two at a time, from *m/z* 1107. Similarly, the signals at *m/z* 966, 867, 740, 669, 556, 429 and 344 can be attributed to the successive loss of one residue from the ion at *m/z* 1051. In the spectrum, the masses at *m/z* 184, 172, 127, 113, 101, 85 and 71 implied the amino acid residues valine+N-methylalanine (or N-methyl-β-alanine) or leucine (or isoleucine)+alanine (or β-alanine), valine+methoxyacetate (73 mu), N-

methylleucine (or N-methylisoleucine), threonine, N-methylalanine (or N-methyl-β-alanine) and alanine (or β-alanine) respectively. Combining these two sets of information with the amino acid analysis result, the peptide fragment **51a** can be mapped out. This assumes that these fragments were formed by the loss of individual amino acid residues one by one from the $(M+H)^+$ ion.



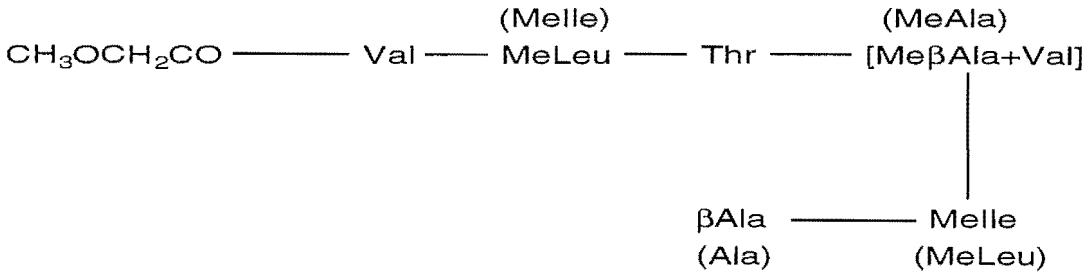
Fragment **51a** of the ring-opened peptide (**51**)

N.B. The amino acids in the parenthesis () may exchange with their isobaric residues.

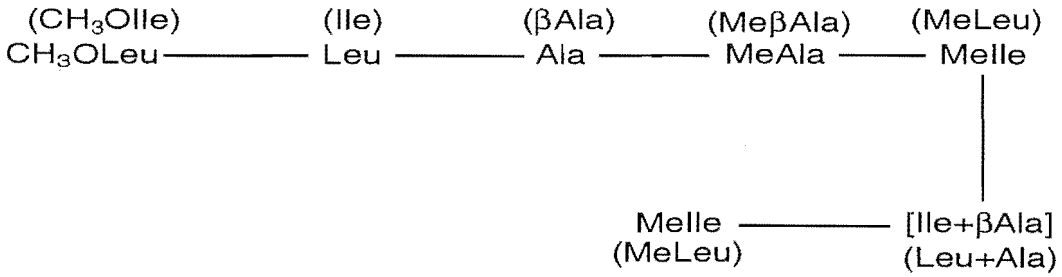
This temporary assignment needed further confirmation. The observed ions could have been generated from internal fragmentation or from different precursor ions. Furthermore, interfering peaks could have arisen from the matrix. In addition, some of the signals were very weak, which added more uncertainties to the assignments. To circumvent the uncertainties presented by the direct analysis of the FABMS spectrum, MS/MS techniques were used to obtain more information. A series of carefully selected fragments were scanned to establish the relationships between the precursor (parent ion) and the product (daughter ion) ions.

Initially, the $(M+H)^+$ at *m/z* 1451 was selected for tandem mass spectrometric analysis (Fig. 4.4.3). The fragmentation was anticipated to occur from both ends of the peptide. From one direction (N-terminus),

the ions at *m/z* 1451, 1279, 1152, 1051, 867, 740 and 556 were recognised. A second series of fragments including *m/z* 1451, 1291, 1178, 1107, 1022, 895, 711 and 584 were traced from the C-terminus of the peptide chain. The peptide fragments **51b** and **51c** were mapped out based on these data.



Peptide fragment **51b**



Peptide fragment **51c**

N.B. The residues in parentheses [] were calculated from the difference of amu.

The link scans yielded rich structural information on fragmentations in the high mass range, which was where the intensity of the signals from a normal scan are usually too low to be distinguished. Comparison of the peptide fragments **51b** and **51c** showed that the whole sequence of **51b** and a part of **51c** was overlapped with sequence **51a**. There was a tripeptide segment with a mass of 342 in **51c** which was missed in **51a**. This tripeptide fitted the "mass gap" between **51a** and the entire linear peptide.

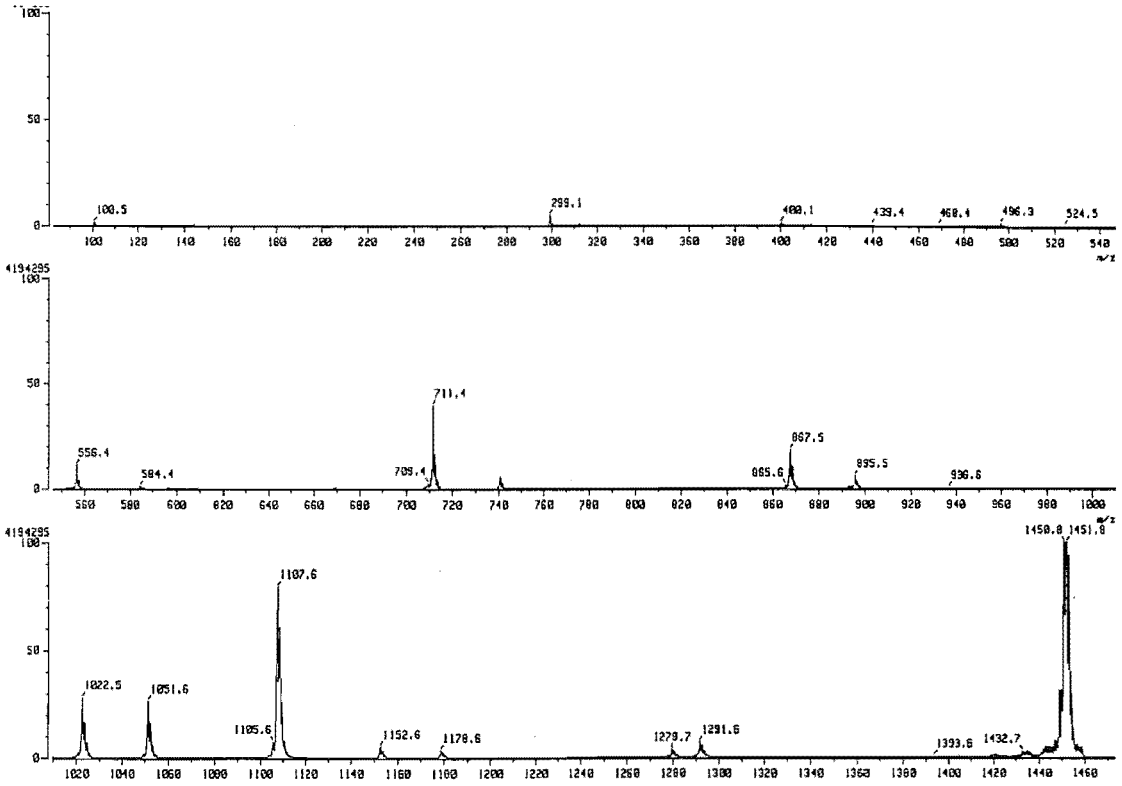


Fig. 4.4.3 Link scan for *m/z* 1451

The ion at *m/z* 1107, one amu less than the ion generated by subtracting 342 from the $(M+H)^+$ ion, was attributed to either a cleavage accompanied by a one hydrogen transfer to the tripeptide part, or the C-terminus carrying an additional H^+ for protonation of the peptide molecule. Re-examination of Fig. 4.4.3 revealed that the mass gap corresponding to the loss of the first amino acid residue from the C-terminus was 159, rather than 158. This corresponded to the fragment from N-methylleucine methyl ester. Since signals from the transfer of one or two hydrogens to the charged fragment are normally more abundant, the above observation strongly suggested that the N-methylleucine at C-terminus carried one additional H.

The verification of the composition of the tripeptide was accomplished by a link for the ion at *m/z* 344 (Fig. 4.4.4). The resultant daughter ions

included *m/z* 312, 284, 185 and 72, which can be reasonably explained by 344-CH_3OH, followed by loss of CO, then 344-N-methylleucine methyl ester-leucine to release an alanine residue (*m/z* 72). The signals at *m/z* 44, 86 and 100 (Fig. 4.4.4) corresponded to the immonium ions (R-CH=NH^+) of alanine, leucine and N-methylleucine. This offered more evidence for the existence of these amino acids. The ions at *m/z* 159 and 160 can be assigned to Y_1' and Y_1'' ions (cf. Fig 4.4.1), while the *m/z* 216 ion probably arose from eliminating $COOCH_3$ at the C-terminus and an alanine residue at the other end of the tripeptide accompanied by a two hydrogen transfer.

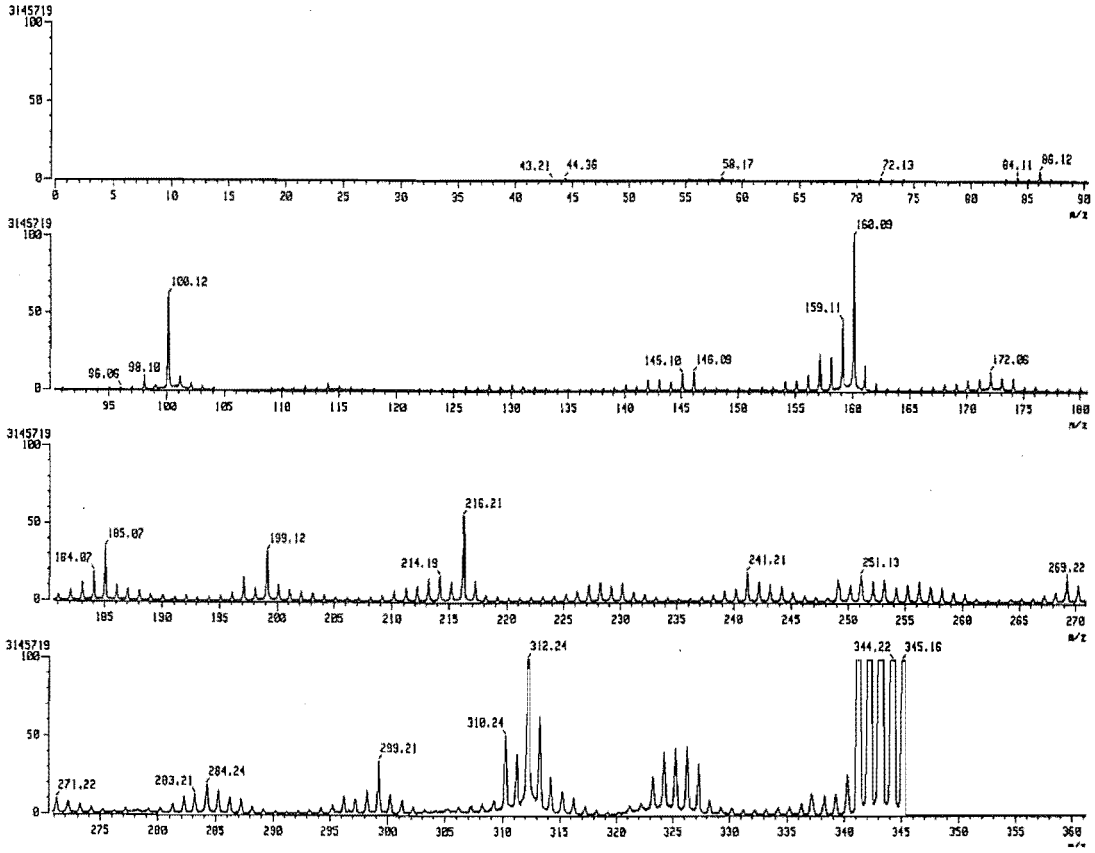


Fig. 4.4.4 Link scan for *m/z* 344

A link scan for *m/z* 429 (Fig. 4.4.5) gave only a weak signal at *m/z* 344, but successive losses of N-methylleucine methyl ester (*m/z* 270), leucine (*m/z* 157) and alanine residues were observed to yield methylalanine (*m/z*

86) that can be recognised along with all the other immonium ions. Therefore, a tetrapeptide fragment was drawn out from the above assignments. The tetrapeptide structure not only makes up the missed C-terminus in **51a**, but also shows one residue overlapping with it.

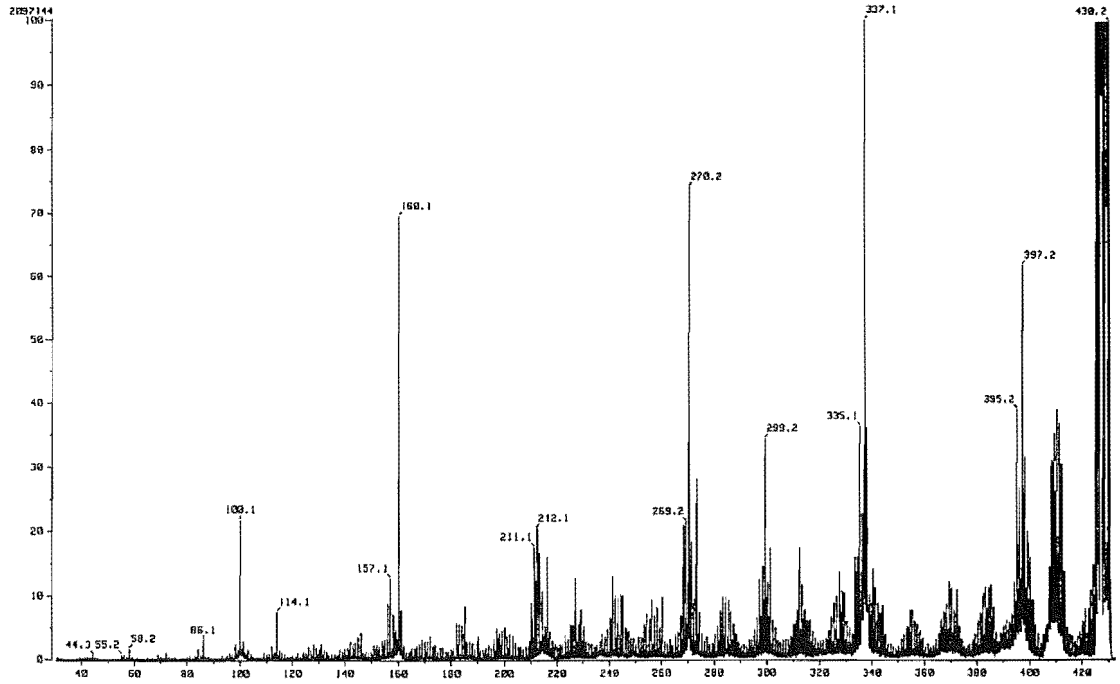


Fig. 4.4.5 Link scan for *m/z* 429

The sequence at the N-terminus was confirmed by the following link scans. A link scan for the ion at *m/z* 299 (Fig. 4.4.6) gave rise to daughter ions at *m/z* 172 (strong) and 100. The ion at *m/z* 172 was assignable to B_1+$COCH_2OCH_3$ (cf. Fig. 4.4.1), which loses the N-terminal blocking group ($COCH_2OCH_3$) to release a valine residue (plus one hydrogen) at *m/z* 100. The elimination of the N-terminal blocking group (72 mu) from *m/z* 299 was also rationalised by observation of a peak at *m/z* 227. The peaks at *m/z* 255 and 144 can be accounted for by *m/z* 299-C_2H_6O (from methoxy acetate) and 172-CO respectively. The fragment *m/z* 299 was also recognised as a daughter ion in a link from *m/z* 400 (Fig. 4.4.7). The peaks at *m/z* 270 and 229, found in the same spectrum, were assignable to A_2 (cf. Fig 4.4.1) and Thr-MeLeu (or Thr-

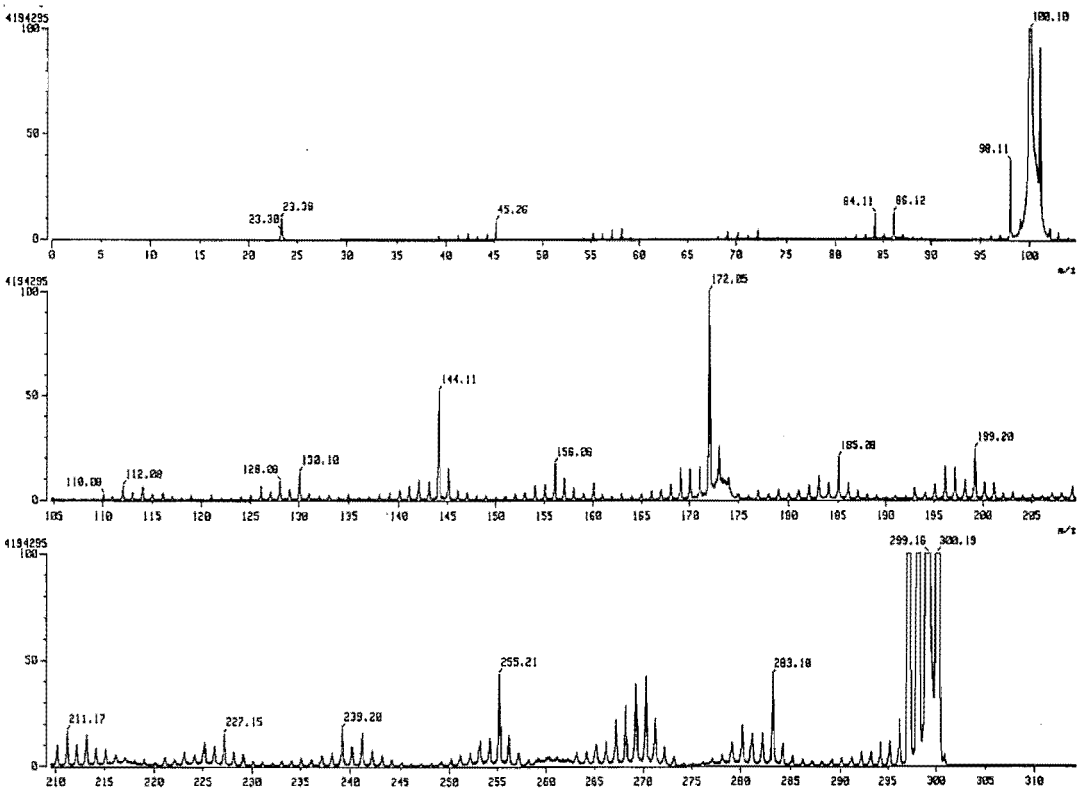


Fig. 4.4.6 Link scan for *m/z* 299

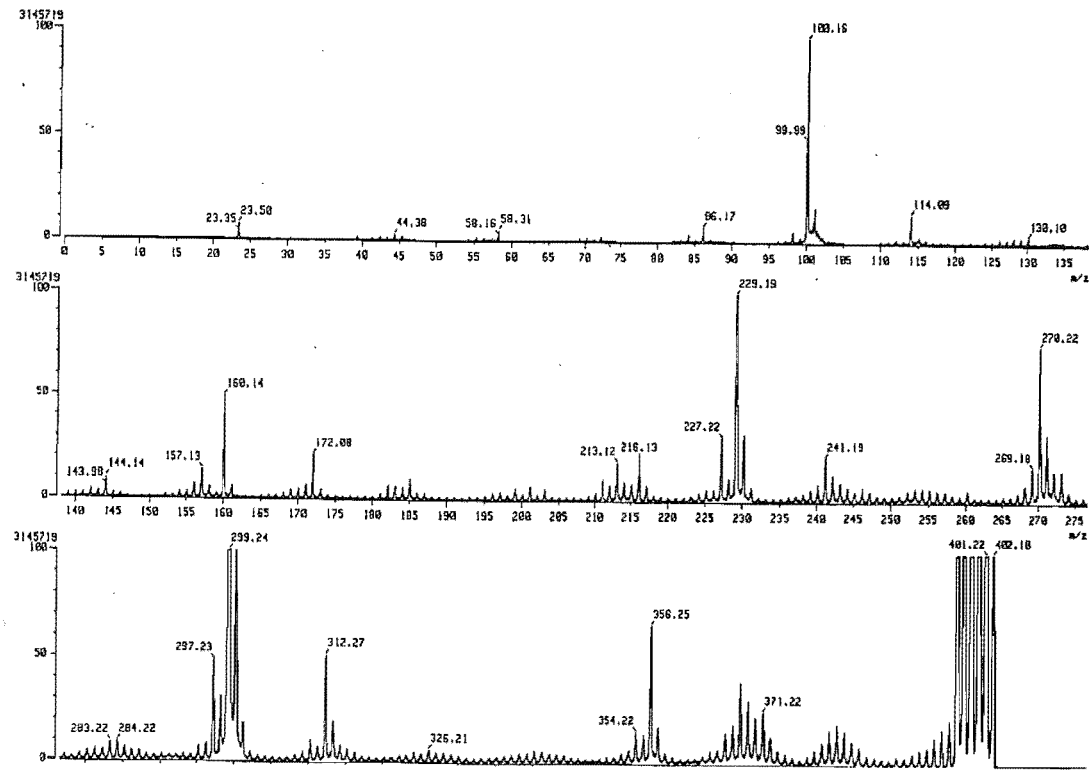


Fig. 4.4.7 Link scan for *m/z* 400

MeIle) fragments. These each confirmed that the threonine was the third residue from the N-terminus. Considering that the threonine was the only hydroxy amino acid in the component amino acids of IIIe, it must form an ester bond with the carboxyl group of MeLeu at the C-terminus to produce a cyclic structure. From this, one can conclude that the side chain for the cyclic depsipeptide can be figured as MeLeu(MeIle)-Val-$COCH_2OCH_3$.

The remaining sequence between the threonine and N-methylalanine was clarified by the link experiments described below. A link for *m/z* 1107 showed that the fragments *m/z* 1022, 895, 711, 584, 400 and 299 were formed from it, but other important peaks at *m/z* 782, 485 and 172, expected in structure of **51a**, were of low abundance. Fortunately, a link for *m/z* 1022 picked up *m/z* 895, 782 and 711 ions at a reasonable intensity, although the intensity of *m/z* 485 was still very low. However, a link from *m/z* 711 gave a reasonably intense peak at *m/z* 485. Other daughter ions at *m/z* 584, 400, 299, 172 and 100 in this link confirmed the existence of the major fragment, MeLeu-Val-MeβAla-Thr-MeLeu-Val-CO-CH_2OCH_3, in the structure of **51a**.

Fragmentations from the other end of the big fragment were also observed.

$$711 \xrightarrow[\text{Val-OAcOMe}]{-171} 540 \xrightarrow[\text{MeLeu}]{-127} 413 \xrightarrow[\text{Thr}]{-101} 312 \xrightarrow[\text{MeIle}]{-127} 185 \xrightarrow[\beta\text{-MeAla}]{-85} 100$$

These fragmentations corresponded to the loss of Val-$COCH_2OCH_3$, MeLeu and Thr from the above fragment, then loss of MeIle from the other end. A link from *m/z* 584 produced ions at *m/z* 485 (weak), 400, 299, 185 and 100, which verified that *m/z* 485 was the daughter ion from *m/z* 584.

The same sequence was obtained by interpretation of the MS/MS spectra originating from the C-terminus fragmentations. Peaks at *m/z* 966 and 669, which were expected in the fragmentation of **51b**, did not show up in the spectrum of $(M+H)^+$. However, *m/z* 966 was recognised as a daughter ion of 1051 in a link scan from *m/z* 1051. Other important daughter ions of 1051, such as 867, 740 and 556 corresponding to Y_8, Y_7 and Y_5 respectively (cf. Fig. 4.4.1), were also found in the same link. Again, there was no peak at *m/z* 669. In order to find the sequence of the dipeptide that comprised the sixth and seventh amino acids from C-terminus, the ions at *m/z* 867 and 740 were scanned. Only a weak peak at *m/z* 669 was found in the spectrum of *m/z* 740. However, an ion at *m/z* 556 was readily recognised as the daughter for *m/z* 867, 740 and 669. These results provided indirect evidence for the assignment of leucine or isoleucine in the sixth and alanine or β-alanine in the seventh position of the linear peptide (from the C-terminus). A strong peak at *m/z* 397, which arose from the loss of a tripeptide from the C-terminus, was observed in the spectrum of the ion at *m/z* 740. This implied that cleavage of the bond of Ile (or Leu)-β-Ala (or Ala) was not favourable. The same phenomenon was also observed when the peptide chain was sequenced from the N-terminus end. The ion at *m/z* 429 was clearly identified as a daughter ion of *m/z* 556 in the link scan for *m/z* 556. This suggested the loss of an N-methylleucine or N-methylisoleucine. The sequencing from *m/z* 1051, corresponding to the fourth residue in from the N-terminus, was accomplished.

The entire sequence of the linear peptide can be assigned by putting all the above information together (see Fig. 4.4.8). The numbers on the structure represent the B_i and Y_i (i=1-13) ions, corresponding to sequencing from the C- and N-termini respectively.

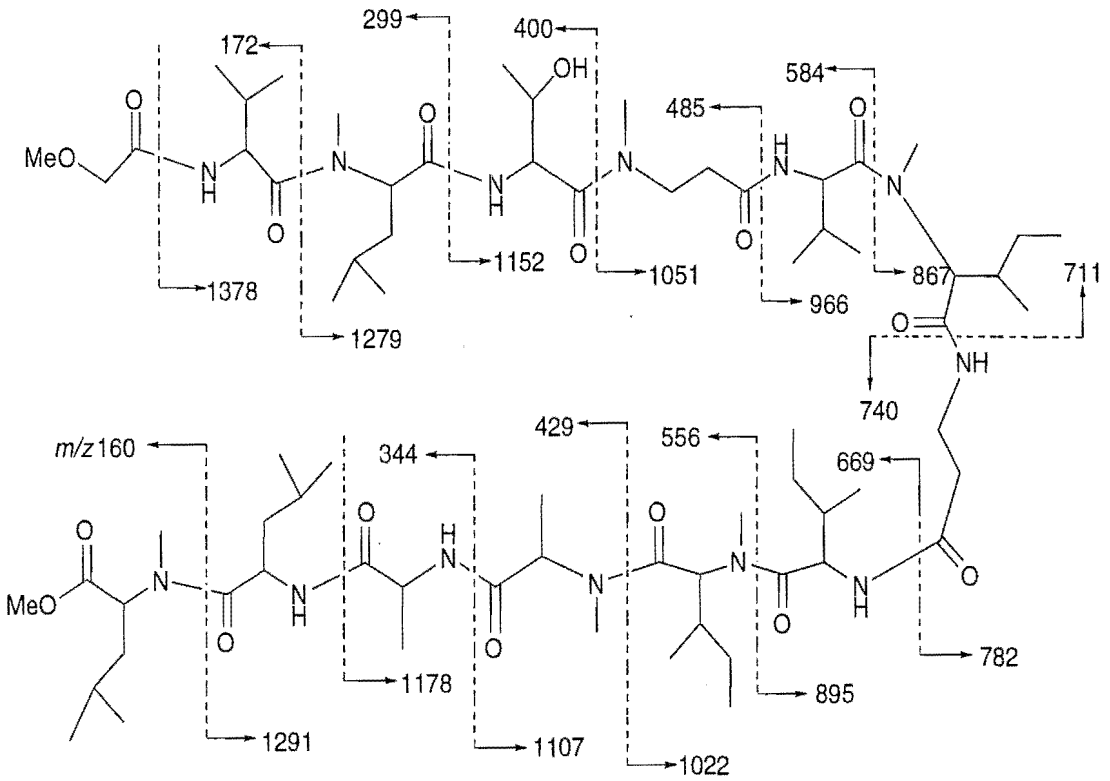
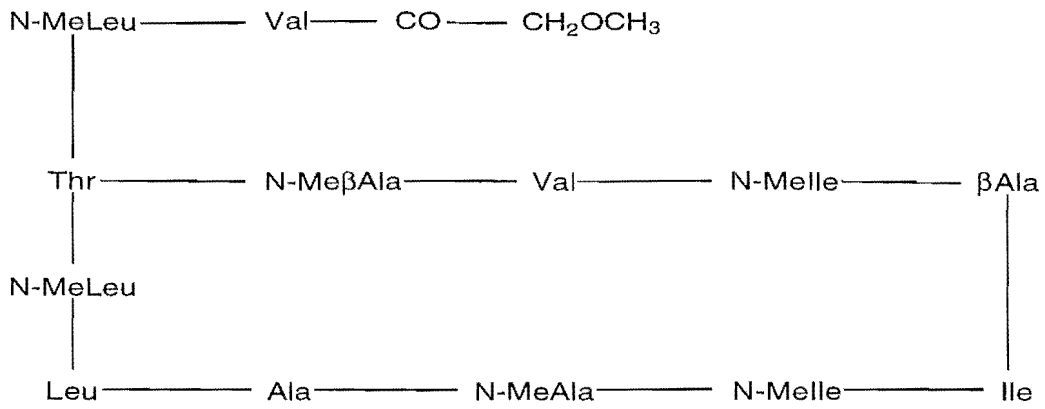


Fig. 4.4.8 FABMS/MS fragmentation ions from both N- and C-terminus of the ring-opened peptide

The cyclic structure of the peptolide was established by joining the hydroxy group of threonine with the carboxyl of the C-terminal N-methylleucine, leading to the assignment of the peptolide as **50**.

However, the positions of the N-methylleucines, N-methylisoleucines, leucine, isoleucine, alanine, β-alanine, N-methylalanine and N-methyl-β-alanine may be replaced by their isobaric isomers respectively. To solve this problem, it was necessary to utilise NMR techniques



The entire structure of theonellapeptolide IIIe (**50**)

4.5 Structural Elucidation by NMR

4.5.1 Strategy for Structure Elucidation by NMR

At this juncture, a tentative structure had been proposed for theonellapeptolide IIIe. This was based on both amino acid analysis and sequencing of the ring-opened peptide. However, the position of some residues was still uncertain. This problem was solved by a detailed analysis of the various NMR spectra. Extensive NMR techniques, including ^{1}H, ^{1}H-^{1}H decoupling, ^{13}C, DEPT, COSY, TOCSY, HSMQC, HMBC and ROESY, using different deuterium NMR solvents, were utilised to determine this complex structure.

The assignment of the individual signals in the NMR spectra were assisted by the COSY and TOCSY data. Considering the relatively large number of carbon atoms and the complexity of the structure, overlaps were expected in ^{13}C NMR spectrum. In the end, carbon spectra were acquired in different solvents and the spectra were then compared in order to resolve the ambiguities. The DEPT (both 90° and 135°) spectra also provided useful information for this purpose, as well as determining the multiplicity of the carbons. The individual spin systems were recognised based on careful interpretation of COSY, TOCSY, HSMQC, HMBC, ROESY as well as ^{1}H-^{1}H decoupling spectra. This allowed the assignment of most proton signals. The linkages between the individual residues were established through two and three bond H-C correlations observed in the HMBC spectra, while observation of the nOe between protons in adjacent residues in the ROESY spectra provided more connectivity data.

4.5.2 Optimising the Conditions for NMR Measurement

The molecular formula, determined by HRFABMS, indicated that theonellapeptolide IIIe contained 127 hydrogen atoms and 71 carbons. So, heavily crowded ^{1}H and ^{13}C NMR spectra were expected. Therefore, high quality ^{1}H and ^{13}C NMR spectra were necessary for the unambiguous assignment of the structure. To achieve the best result, the optimisation of the NMR solvent was carried out by either changing the type of the deuterium solvent, or varying the ratio of the component solvents where solvent mixtures were used. Temperature effects on the overlap problem were also examined.

Initially, deuterochloroform was used to acquire the ^{1}H NMR spectrum. Seven peaks between δ 6.7-8.8 were recognised, although a pair of peaks exhibited partial overlap due to line broadening. Five sharp singlets were found between δ 2.6-3.4. A ^{13}C spectrum was acquired at 23°C, and 11 signals were detected in the carbonyl carbon area (δ 168-176) with two split peaks. The split peaks were presumed to arise as a consequence of the slow interchange between different conformations. When the probe temperature was increased to 45°C, the splitting either collapsed or decreased. Most of the peaks became sharper. This observation supported the assumption of slow interchanging between conformations. However, only 11 carbonyl resonances were distinguished rather than the expected 14.

When 13% CD_3OD in $CDCl_3$ was used as the NMR solvent, the resolution in the N- and O-methyl regions was improved and a further methyl singlet was recognised. This improvement suggested that CD_3OD was a good modifier for this compound. Next, 100% CD_3OD was tested. As expected, the resultant spectrum was much better than those obtained

previously. In the downfield portion of the spectrum, all seven signals were well resolved as doublets. Seven methyl groups between δ 2.6 and 3.4 were observed as very sharp singlets. At least seven pairs of doublets in the upfield range (δ 0.78-1.44), as well as several other groups of isolated signals, were clearly defined. CD_3OD was an excellent solvent for the proton spectrum. A similar improvement was observed in the ^{13}C spectra using this solvent. Fourteen signals were found between δ 170-177. Sixty seven out of the total 71 carbons were distinguishable. Most of them were also quite sharp. Unfortunately, slow exchange occurred between some of the exchangeable protons in the low field area and CD_3OD over a period of less than 90 minutes. Even after adding 10%-20% of $CDCl_3$, the intensity of the exchangeable proton peaks still decreased quite obviously within two hours. This established that CD_3OD was not a suitable solvent for running more complex experiments, such as HMBC and ROESY, which usually require longer acquisition times. However, these spectra were still worthwhile as references to sort out overlaps in the spectra acquired in other solvents.

To avoid the exchangeable proton problem, DMSO and d_6-pyridine were examined as alternative solvents. In the DMSO spectrum, the peak patterns at low field became more complex. At the same time, peak-broadening was remarkable in the other parts of the spectral range. These observations implied that the interchange among the conformations of the cyclic peptide was slowed in DMSO. For this reason, DMSO was not a good solvent to use. In contrast, the signals in most parts of the spectrum in pyridine-d_6 were much sharper than those in DMSO. Three drops of H_2O was used as a modifier in the pyridine. The resultant spectrum showed an improvement in the low field portion, but the

modification caused line broadening in other parts of the spectrum. It was concluded that this solvent was not worth modifying further.

A review of the literature resulted in a reconsideration of the use of d_4-methanol. A low percentage of H_2O in CD_3OD had frequently been used to acquire NMR spectra for proteins and polypeptides.[153] In the case of theonellapeptolide IIIe, 10% H_2O was used to suppress the deuterium exchange. The spectrum was then re-measured at six hour intervals. It was found that the peaks corresponding to the exchangeable protons were still distinguishable in the spectrum obtained after twelve hours at room temperature, and that the resolution of the spectrum was comparable to that obtained in CD_3OD. The drawback with this choice of solvent was the strong H_2O peak that appeared at δ 4.7. This resonance disturbed other important resonances around it. These resonances corresponded to the α-protons of the amino acids.

The last group of solvents tested for optimisation purposes was a mixture of d_6-benzene and CD_3OD. This pair was selected on the assumption that methanol could improve the resolution of the signals, while addition of C_6D_6 would dilute the concentration of the exchangeable deuterium atoms in the solvent, effectively reducing the rate of exchange. With a 1:1 mixture of CD_3OD/C_6D_6, the results looked very encouraging. The resolution was still very satisfactory, especially in the low field portion, and the methyl resonance area. A series of spectra were next obtained 3, 6, 12 and 72 hours after the sample had been prepared in this solvent system. The probe temperature was maintained at room temperature. There was no obvious decrease in the intensity of the active proton signals in the spectrum after three hours. The intensity of the peaks at δ 7.3 and 7.5 decreased gradually after six hours. These two signals had

disappeared after 72 hours. When the ratio of CD_3OD was reduced to 40%, the signal at δ 8.78 had diminished discernibly after four hours standing. A decrease in the rate of exchange was observed when the ratio was changed to 30% CD_3OD/C_6D_6. At this ratio, the resolution was still good, except that the two methyl signals around δ 3.2 were less well resolved. A further decrease in the percentage of methanol to 20% resulted in a further improvement of the methyl resonances, but caused a partial overlap for two signals at about δ 3.8 which had been well resolved previously in other solvent mixtures. The spectra were measured in this solvent at 0, 5 and 18 hours after sample preparation. The exchangeable protons lasted for 18 hours without obvious collapse. Reducing the ratio to just 10% CD_3OD/C_6D_6 caused greater overlap for the signals, while pure C_6D_6 turned out to be worse than the mixed solvent. After assessment of the data, a mixture containing 20% CD_3OD and 80% C_6D_6 was chosen as the NMR solvent for further spectral studies. This represented a balance between resolution and the lifetime of the exchangeable protons. However, if the sample stayed in the NMR tube for more than ten hours, it needed to be re-converted to the fully protonated form. This was accomplished by standing the sample in CH_3OH for more than six hours.

4.5.3 Characteristics of the ^1H NMR Spectrum

There were seven doublets in the down-field region of the ^1H NMR spectrum (see Fig. 4.5.1). The peak at δ 9.0 was broadened, but was clearly a doublet in the spectra with CD_3OD or 10% H_2O/CD_3OD as the solvent. These signals were assigned to seven amide protons. Between δ 4.9 and δ 5.5, several multiplets heavily overlapped each other. From the integral values this group of peaks consisted of ten protons. They were

attributed to the α-protons of the amino acid residues or the β-proton of threonine. The multiplicity can be accounted for by coupling from either the β-protons on the side chain or both the β- and amide protons. The resonances at δ 4.0-5.0 were due to the α-H in α-amino acids or the β-H in β-amino acids. A two proton singlet at δ 3.86 (which appeared as a doublet in the CD_3OD spectrum) indicated the existence of an isolated methylene group. This is a characteristic peak for methoxyacetate. The broad peaks at δ 3.53 and δ 3.73 were assigned to either α- or β-amino acids. The most distinctive characteristic features of this spectrum were the seven sharp singlets at δ 3.33, 3.30, 3.24, 3.19, 3.16, 2.75 and 2.69. These signals are typical of methyl groups attached to nitrogen or oxygen atoms. Since only six N-methyl amino acids were identified from the amino acid analysis, the seventh must be a methoxy group. At higher field in the spectrum, six pairs of doublets stand above complex multiplets. Normally, this is the aliphatic methyl area. The side chains of alanine, N-methylalanine, valine, leucine, N-methylleucine, isoleucine, N-methylisoleucine and threonine are responsible for these doublets. The remaining 65 hydrogen atoms were crowded into the δ 0.8 to 2.6 range.

The ^{13}C NMR spectrum obtained in 20% CD_3OD/C_6D_6 showed 13 signals between δ 169 and 176, while 14 were visible in CD_3OD. A comparison of these two spectra revealed that the peak at δ 172.5 was comprised of two carbon signals. These peaks arose from the carbonyl groups of the peptide bonds. The 14 carbonyl signals can be assigned to the 13 amino acids, and a methoxyacetate. Fifty two of the remaining 61 carbons lay between δ 9-72. Of these, the resonances at δ 72.1 and 71.2 were assigned to the carbons attached to oxygen. The DEPT data further distinguished the former resonance as CH_2-O and the latter as being due to CH-O. Since only one CH_2 and one CH group were directly connected

to an ether and an ester oxygen respectively, the δ 72.1 signal was assigned to the β-carbon of methoxyacetate, while δ 71.2 was assigned to the β-carbon of threonine. Ten resonances, the α-carbons of α-amino acids or the methoxyl carbon, were found between δ 52 and 63. However, the spectrum measured in 10% H_2O/CD_3OD yielded 11 signals in that region, so at least two carbons were overlapped in 20% CD_3OD/C_6D_6. The DEPT spectrum confirmed that the resonance at δ 59.0 arose from a methyl group which was assigned to the methoxyl of methoxyacetate. Furthermore, the signals at δ 45.4, 40.8, 38.8, 38.3, 26.9, 25.6 and 25.3 arose from the resonances of CH_2 groups. The other resonances could not be readily assigned at this stage.

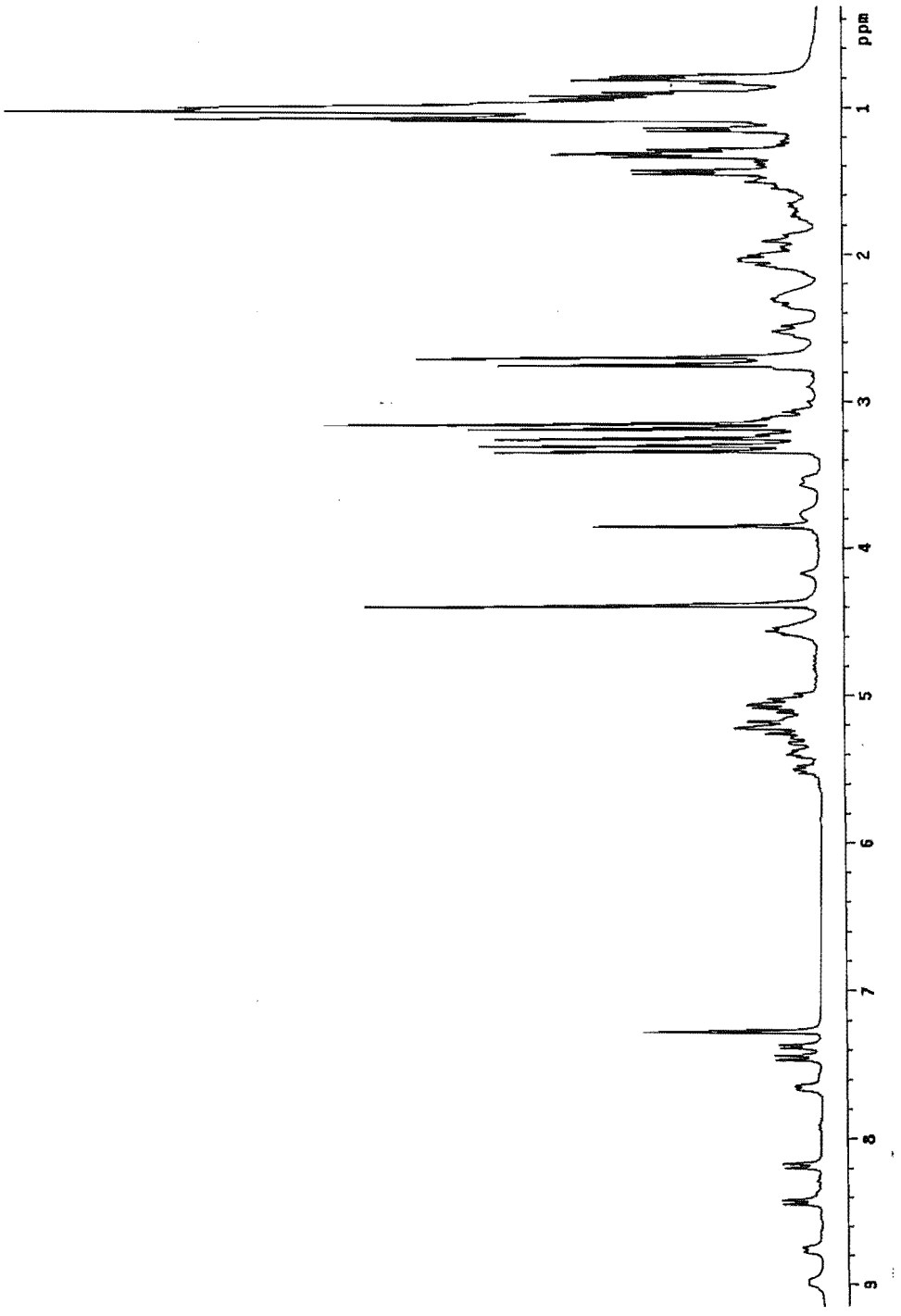


Fig. 4.5.1 ^{1}H NMR spectrum of theonellapeptolide IIIe (**50**)

4.5.4 Assignment of the Spin Systems

Theonellapeptolide IIIe consisted of 13 amino acids. Therefore, there should be at least the same number of spin systems present in the ^{1}H NMR spectrum. The spin systems of the seven amino acid residues with an amide hydrogen were traced from the amide proton, through the corresponding residue, by 2D TOCSY. Those with N-methyl groups turned out to be troublesome. It was necessary to examine all possible correlations and make careful comparisons with each spin system.

In the ^{1}H NMR spectrum of IIIe, the most distinctive signal was a well isolated sharp singlet at δ 3.86. Unambiguous assignment of this signal was used as a firm base for further extension of the structure through the spectral analysis. Fortunately, the signal of its attached carbon (δ_C 72.1) was also readily recognisable in the HSMQC spectrum. Similarly, the methoxy proton signal (δ 3.16) was located by looking at the cross peak from the carbon signal at δ 59.0. The HMBC spectrum revealed correlations from the methoxy protons to a methylene carbon, and from the methylene protons to the methoxy carbon, as well as to a carbonyl carbon (δ 170.4). Thus, the methoxyacetate moiety was confirmed. Because the methylene proton showed cross peaks to the methoxy and an amide resonance at δ 7.45 in the ROESY spectrum, this partial structure was verified.

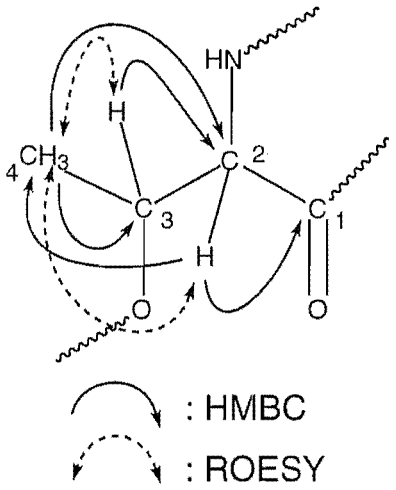


Fig. 4.5.2 HMBC and ROESY correlations of threonine residue

The 2D TOCSY spectrum indicated that the amide proton at δ 9.0 was coupled to the signals at δ 5.47, 5.0 and 1.31. In turn, the HSMQC spectrum revealed that the proton at δ 5.47 was attached to the carbon at δ 71.2, assigned to the β-carbon of threonine by the DEPT spectrum. Hence, the β-H of the threonine was located. The proton at δ 5.0, attached to the carbon at δ 53.7 (HSMQC), can be placed in the α-position. The resonance at δ 1.31 was also observed as a sharp doublet, and exhibited correlations to both of the α- and β-carbons in the HMBC. The above data suggested a threonine residue (see Fig. 4.5.2). This segment was confirmed from the following correlations observed in the ROESY spectrum: NH/H2, NH/H3, H4/H3 and H4/H2.

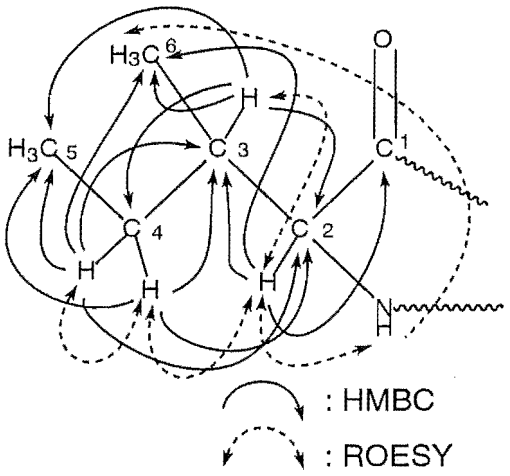


Fig. 4.5.3 HMBC and ROESY correlations of *allo*-isoleucine residue

The amide proton at δ 8.76 was coupled to the resonances at δ 5.38, 1.75, 1.38, 1.04 and 0.98 (2D TOCSY). Further examination of the TOCSY spectrum resulted in the observation of correlation peaks at δ 8.76, 1.75, 1.55 (weak), 1.30, 1.18 and 0.98 from the δ 5.38 signal. In the COSY spectrum, the proton at δ 8.76 was coupled to a proton at δ 5.38, then that at δ 5.38 to δ 1.75, and then that at δ 1.75 to δ 1.22. The HSMQC spectrum showed that the δ 5.38 and 1.75 signals were methine protons, while the two protons at δ 1.55 and 1.22 attached to the same carbon at δ 26.9. The HMBC spectrum showed the following cross peaks: H2/C3 and C4, H3/C2, C4, C5 and C6, H5/C3, C4 and C6 (see Fig. 4.5.3). Combining the above evidence yielded an isoleucine moiety.

This partial structure was further defined by the ROESY spectrum (Figure 4.5.3). Observation of the correlation between NH and the methyl group at δ 0.98 implied that these two groups were close to each other in space. The correlations from H3 (δ 1.75) to H2, H4 and H6 are indicative of its location at the β-position of isoleucine. The two methylene protons, correlated to each other, showed correlations to H2. So, the second residue was identified as isoleucine.

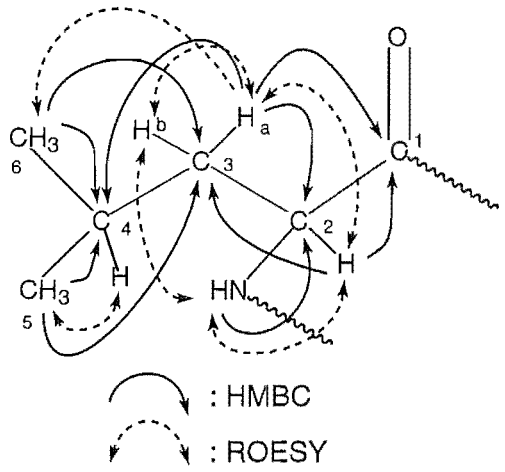


Fig. 4.5.4 HMBC and ROESY correlations of leucine residue

The TOCSY spectrum showed that the third amide signal (δ 8.43) was coupled with resonances at δ 5.15, 1.87, 1.34 and 0.99, while the proton at δ 1.87 correlated with δ 8.43, 5.15, 1.34, 1.04 and 0.98. The latter set confirmed that these signals all belonged to the same spin system. The HSMQC experiment revealed that the protons at δ 1.87 and 1.34 both correlated to a carbon at δ 40.8. This was confirmed as a secondary carbon by the DEPT spectrum.

The HSMQC and DEPT spectrum identified two methine groups, δ 1.86/25.4 and 5.15/48.8. The latter was assigned as the α CH. Coupling observed between the protons at δ 5.15 and 1.87, and 1.87 and 1.34 in the COSY spectrum suggested that δ 1.87, and by implication the δ 1.34 H, were related β-protons.

An HMBC experiment provided more correlations between NH and C1, H2 and C3, H3 (δ_H 1.87) and C2, H3 and C4 (δ_C 25.4), as well as from both H5 (δ_H 1.04) and H6 (δ_H 0.98) of the terminal methyl groups to C3 and C4 respectively. These data supported the assignment of the methine at δ 1.86/25.4 to the γ position of the amino acid. H2 exhibited a two

bond correlation to C3 at δ 40.8. By gathering the information from these analyses together, a leucine fragment could be defined (Fig. 4.5.4).

A ROESY experiment, which displayed a correlation between H4 and H5 at δ 1.86, added more assurance to the above assignment. Other correlations observed in the ROESY spectrum included NH/H2, H2/H3a, NH/H3b, H3a (δ 1.87)/H3b (δ 1.34) and H3/H6. All these observations are consistent with the structure of leucine.

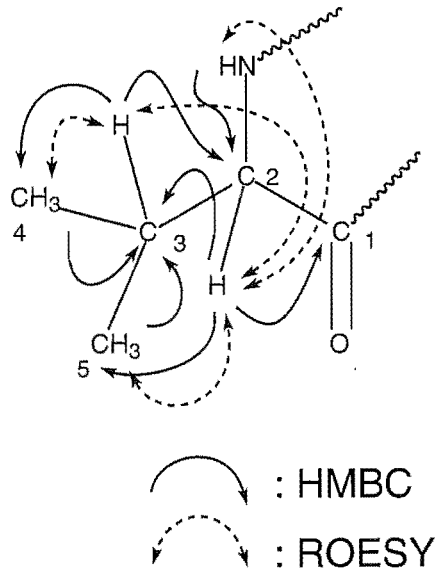


Fig. 4.5.5 HMBC and ROESY correlations of valine (2) residue

Correlation peaks from the NH at δ 8.19 to protons at δ 5.06, 2.06 and 0.99 were observed in the 2D TOCSY spectrum. The COSY and HSMQC spectra revealed that the NH proton at δ 8.19 was coupled to a methine (δ_H 5.06/δ_C 55.4), which in turn was further coupled to another methine (δ_H 2.06/δ_C 31.0) to which two methyl groups (δ_H 0.98/δ_C 18.3 and δ_H 0.99/δ_C 19.7) were attached. A valine unit was deduced: it was assumed that these five groups of signals represented all the protons in the spin system. To obtain confirmatory evidence for this assumption, a trace at δ 5.06 in the TOCSY spectrum was examined. A new amide peak at δ 7.45 was observed along with the expected δ 8.19 signal, as well as other odd

resonances. This result can not serve as confirmation, but did imply that another spin system contained a proton resonating at around δ 5.06.

More evidence for the valine assignment was extracted from the HMBC spectrum. Correlations from the NH to C2 (first methine), H2 to C3 (second methine) and the carbonyl carbon at δ 174.0, H_3 to C2 and C4, and the two methyl protons to C3 provided enough information to confirm valine (Fig. 4.5.5).

If further substantiation was necessary, this was provided by the observation of cross-peaks in the ROESY, such as NH/H2, H2/H3, H2/H4, H2/H5, H3/H4 and H3/H5.

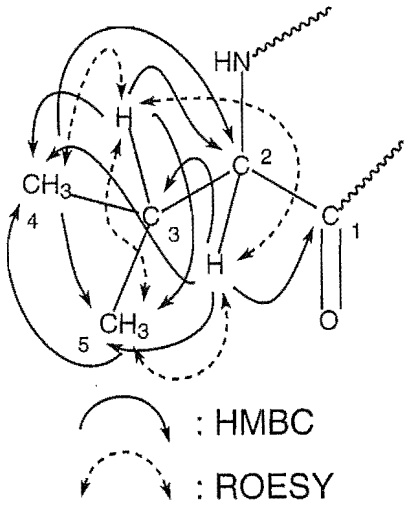


Fig. 4.5.6 HMBC and ROESY correlations of valine (1) residue

As noted above, there is more than one resonance at δ 5.06. The trace at δ_H 5.06 in the TOCSY spectrum had indicated correlations with an NH at δ 7.45. From this, the NH signal correlations were observed to δ 5.03, 2.01, 0.98 and 0.91. The HSMQC and DEPT spectra indicated that δ 5.03 and 2.01 were attached to different methine carbons (δ_C 54.5 and δ_C 32.1 respectively); δ 0.98 and 0.91 were connected to primary carbons. In the COSY spectrum, coupling patterns were observed that were similar to the valine residue described previously. As two valines were found in the

amino acid analysis, assignment of this spin system to another valine residue fitted the amino acid analysis data.

HMBC correlations from H2 to C3, and from the protons of the two terminal methyl groups to C3, were observed. Although a correlation from NH to C2 was not found, successive connection of all segments within a valine subunit could still be made through the observation of correlations to the same carbonyl (δ_C 170.4) from both NH and H2.

Because the chemical shifts of most carbons and hydrogens in the two valine residues were very similar, unambiguous assignment of the structures was not straightforward. The proton at δ 5.03 was located by observation of the correlations from it to two carbonyls at δ 174.4 and 170.4. The latter had been assigned to the methoxyacetate. As this segment was the blocking group at the N-terminus of the peptide, the corresponding valine can be assigned as the N-terminal amino acid, hence the residue was named valine 1. The first carbonyl group must therefore belong to valine 1. The proton at δ 5.06 for the other valine clearly correlated with the carbonyl carbons at δ 174.0 and 172.9. This residue was designated as valine 2.

The signals from the four terminal methyl groups of these two valines gave a broad peak around δ 1.0. For valine 1, the starting point chosen was at H3 (δ 2.01). This position had been established by correlation to C2 in the HMBC spectrum and confirmed through the ROESY correlations between H3 and H2 as well as H3 and the N-methyl signal at δ 3.19, which had been identified as a part of an adjacent amino acid residue. H3 also showed correlations to two primary carbons at δ 19.6 and 17.6. In the HSMQC spectrum, these two carbons were found to be connected with the protons at δ 0.98 and 0.91 respectively. Further

evidence for this arrangement was obtained from the correlations between H3 and H4, as well as H5. In addition, the correlations between H2 and H4, and between the N-methyl protons and H4, provided more support. The key correlations in both of the HMBC and ROESY spectra were depicted in figure 4.5.6. Using the same strategy, H3 and the two methyl protons in valine 2 (see Fig. 4.5.5) were determined.

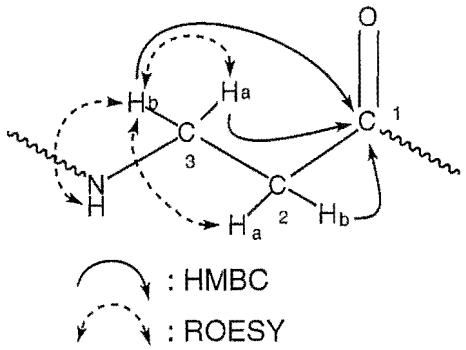


Fig.4.5.7 HMBC and ROESY correlations of β-alanine residue

The spin system of the β-alanine residue NH at δ 7.65 was established by comparing two TOCSY traces which originated from either the NH at δ 7.65 or the δ 3.53 H. It consisted of protons resonating at δ 4.54, 3.53, 2.69 and 2.51. The first two were part of a methylene system (δ 35.8), as were the second pair (δ 35.9, from analysis of HSMQC data). The existence of two adjacent methylene groups was consistent with a β-alanine partial structure. The resonances at δ 4.54 (H3a) and 3.53 (H3b) were assigned to the β-methylene on the basis of chemical shifts. This designation was supported by the HMBC correlations: NH/CO (δ_C 170.8), H3a/CO (δ_C 170.8), H3b/CO (δ_C 170.8), H3a/CO (δ_C 172.5), H3b/CO (δ_C 172.5) and H2/CO (δ_C 172.5). The complete assignment is shown in Fig. 4.5.7.

The last remaining unassigned amide proton (δ 7.39) was attributed to alanine. In the TOCSY spectrum, this resonance correlated with δ 4.53 and 1.29. The H/C relationship was established by HMBC. The HMBC

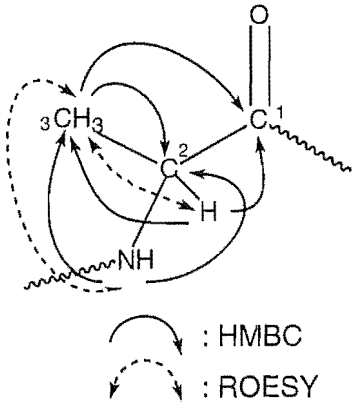


Fig. 4.5.8 HMBC and ROESY correlations of alanine residue

correlations from NH to C2 (δ_C 52.3), and H2 to C3 (δ_C 17.9) provided substantiating evidence for the proposed partial structure (see Fig.4.5.8).

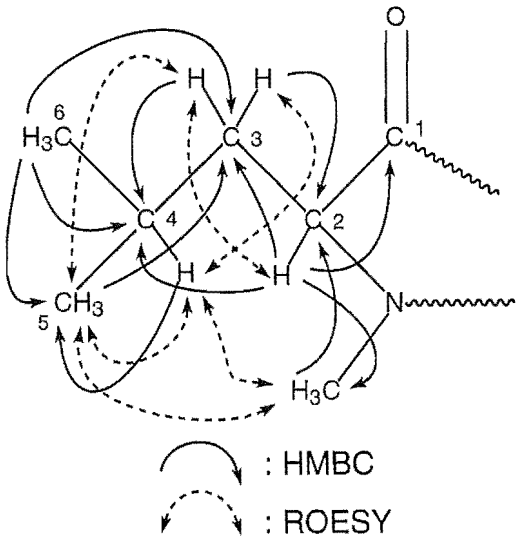


Fig. 4.5.9 HMBC and ROESY correlations of N-methylleucine(1) residue

N-methylleucine (Fig. 4.5.9) was identified as follows: cross peaks from an α-hydrogen at δ 5.28 in the TOCSY spectrum included resonances at δ 2.05, 1.48, 1.06 and 0.94. The absence of an amide signal at lower field in this array was indicative of an N-methyl amino acid. The C-H connections based on one bond C-H correlations in the HSMQC spectra included H (δ 5.28)/C (δ_C 56.4, CH), H (δ 2.05) and H (δ 1.48)/C (δ_C 38.0, CH_2), H (δ 1.48)/C (δ_C 25.9, CH), H (δ 1.06)/C (δ_C 24.1, CH_3) and

H (δ 0.94)/C (δ_C 10.7, CH_3). From these HSMQC data it was clear that the signal at δ 1.48 comprised 2 overlapped signals, as the signal correlated with a methine carbon (δ 25.9) and also with a methylene (δ 38.0) (DEPT). In the HMBC spectrum, this same proton (δ 1.48) also showed correlations to the C2 at δ 56.4. Interpretation of the full set of HMBC data for all protons and carbons in this spin-system (Fig. 4.5.9) allowed the assignment of an N-methylleucine moiety. The assignment of the terminal methyl groups initially relied on HMBC correlations from H4 to C6, then C6 to attached hydrogen atoms by HSMQC spectrum. The chemical shift assignment to H5 was by default after all other the carbons were located.

H2 showed a three bond correlation in the HMBC spectrum to a carbon at δ 31.9, which was found to be connected to a three proton singlet resonating at δ 3.19 (HSMQC). This methyl signal also correlated with the H4 and H5 signals in the ROESY spectrum. These analyses located the N-methyl group for this residue. The other ROESY correlations between H2 and H3, H3 and H4, H3 and H5, and H4 and H5 offered more support for the above assignment (see Fig 4.5.9).

The structures and correlations found for the two N-methylisoleucines are depicted in Figures 4.5.10 and 4.5.11. They were named N-methylisoleucine 1 (MeIle1) and N-methylisoleucine 2 (MeIle2) (counting from the N-terminus of the peptide). Because the majority of the protons in these two residues have similar chemical shifts, the assignments had to be made by frequent comparison of the data from different spectra. For example, the 2D TOCSY trace at δ 5.20 corresponded to the α-proton of both MeIle residues and showed cross peaks corresponding to the protons from *both* residues. In contrast, the

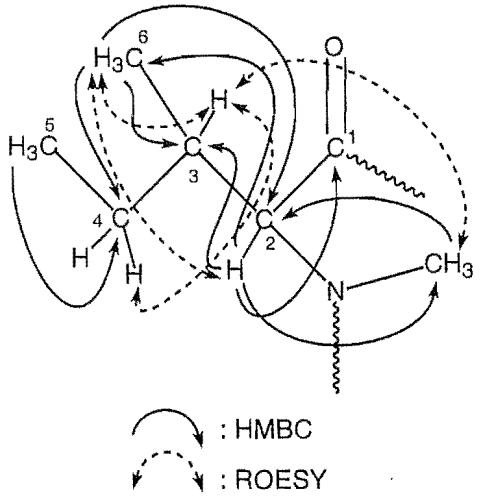


Fig. 4.5.10 HMBC and ROESY correlations of N-methylisoleucine(1) residue

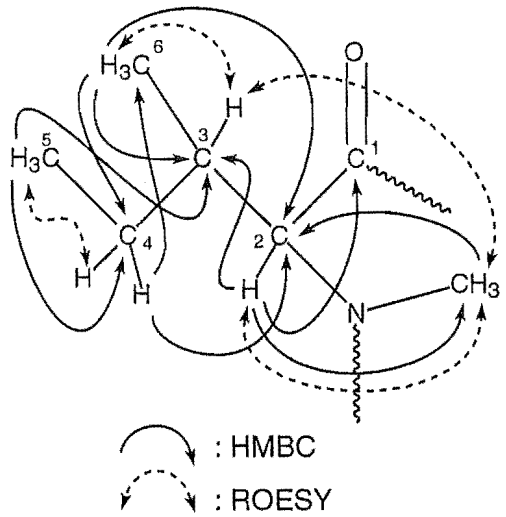


Fig. 4.5.11 HMBC and ROESY correlations of N-methylisoleucine(2) residue

trace for the resonance at δ 1.13 seemed to represent only the MeIle1 spin system. However, the HSMQC spectrum revealed that the C and H chemical shifts for the methylenes and methines in both residues were very close. The proton at δ 1.13 was placed at C6 by observing the HMBC correlation to a carbon at δ 61.4, the α-carbon with its attached proton at δ 5.22 (HSMQC). This observation also established that this

moiety was an N-methylisoleucine residue rather than N-methylleucine. The position of the residue was located as both the H2 of MeIle1 and the NH of β-alanine correlate to the same carbonyl carbon at δ 170.8 in the HMBC spectrum. The other correlations from both H2 and H6 to C3, and from both H6 and H5 to C4 helped to assign the entire moiety. The ROESY spectrum provided the verifying evidences from correlations: H2/H6, H3/H6, H3/NCH_3 (δ 3.24) and H3/H4.

The highest field proton (δ 0.78) in the ^1H NMR spectrum was assigned to C6 of MeIle2 based on the HMBC correlation to C2 at δ 55.5. This was also characteristic of N-methylisoleucine. The HMBC correlations from H2, H6 and H5 to the same carbon (δ 33.8) allowed the location of the C3 position. The methylene proton at δ 1.42 correlated with C6 (δ15.8) which allowed assignment of the C4 protons in MeIle2 residue. The ROESY correlations between H4/H5 and H3/H6 helped to distinguish the two methyl groups.

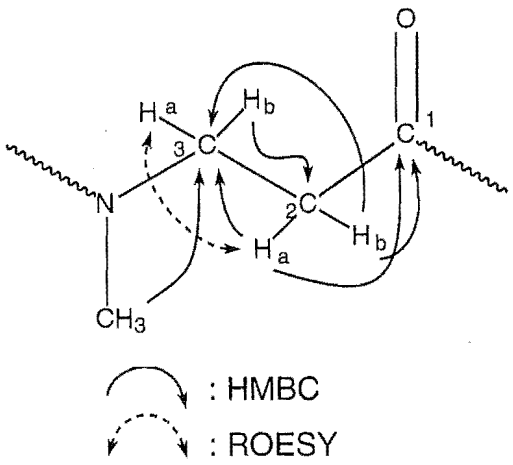


Fig. 4.5.12 HMBC and ROESY correlations of N-methyl-β-alanine residue

N-methyl-β-alanine was the hardest residue to delineate in the 2D TOCSY spectrum, because the signals were mixed with many others. The main signals at δ 4.58, 2.72, 2.51 and 2.30 were picked out from

interfering peaks. This was done by comparison with the TOCSY trace of β-alanine. The correlations in an HSMQC experiment revealed that the δ 4.58 and 2.72 protons were part of a methylene system (δ 45.4), and that another pair, δ 2.51 and 2.30, were attached to another methylene carbon (δ 34.7). The methylene at δ 45.4 was assigned to the β-position based on chemical shift arguments and the correlation from a N-methyl proton at δ 2.75 to δ 45.4 (HMBC spectrum). The big difference in chemical shift between these two protons was probably due to anisotropic shielding from the nearby carbonyl group. This assumes that the proton at δ 2.72 was lying in the shielding area. Correlations from the two α-protons to the C3 and a carbonyl carbon at δ 172.9 were also observed in the HMBC spectrum. The ROESY correlations between these four protons further supported the assignment (see Fig. 4.5.12).

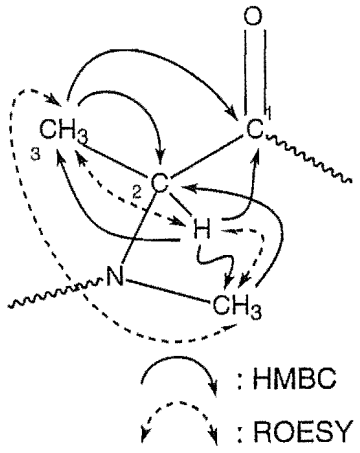


Fig. 4.5.13 HMBC and ROESY correlations of N-methylalanine residue

An attempt to locate the N-methylalanine spin system in the 2D TOCSY spectrum failed as all the signals in this spin system were overlapped with others. A sharp doublet at δ 1.44 remained unidentified after assignment of the other residues. It was presumed to be the signal of the methyl group in N-methyl alanine, due to its relatively low field chemical shift.

Later, a decoupling experiment was utilised to verify this assumption. When the doublet was irradiated, multiplets around δ 5.20 were simplified. The doublet became a sharp singlet if the δ 5.20 signal was irradiated. Even though this was not confirmatory evidence, it definitely established a relationship between these two group of protons.

The HSMQC experiment clearly showed that the proton at δ 1.44 was attached to a primary carbon at δ 15.0. In the HMBC spectrum, this proton correlated to a methine carbon at δ 57.5, to which the proton resonating at δ 5.20 was assigned by HSMQC. This methine proton was in turn correlated to the methyl carbon. This confirmed the N-methyl alanine structure as only these two groups of protons were found in the spin system. The ROESY correlations between the side chain methyl group and H2, as well as with the N-methyl protons at δ 2.69, offered more connectivity information (see Fig. 4.5.13). These ROESY data allowed assignment of the N-methyl group at δ 2.69 to the N-methylalanine residue.

N-methylleucine (2) was another difficult residue. The difficulties were compounded by the low intensities of a resonance at δ 1.66, and two weak signals, compared with the main peaks at δ 2.05, 1.06 and 0.98, in a 2D TOCSY trace from δ 3.74. The H3 (δ 2.05)/C (δ 38.3) correlation in the HSMQC spectrum was hard to distinguish from H3-C3 of N-methylleucine (1), while H2 at δ 3.74 did not show correlations to any carbon atoms in the HMBC spectrum. Assignment of H2, as the α-proton of the amino acid, was in doubt since it appeared at much higher field compared with the other amino acid residues.

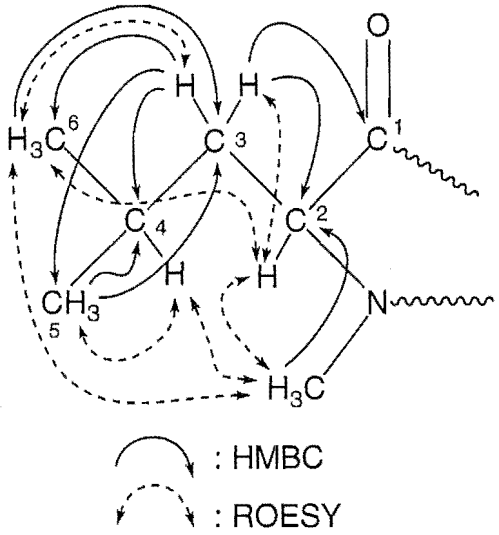


Fig. 4.5.14 HMBC and ROESY correlations of N-methylleucine residue (2)

The COSY spectrum showed that H2 was coupled with H3, but failed to identify any coupling from H3 to H4 (δ 1.66). The linkage between these three groups of protons was deduced from two 1D TOCSY experiments by observing correlations between Ha and Hb, and between Hb and Hc respectively. Examination of the ROESY spectrum resulted in recognition of the following correlations: H2/NCH_3 (δ 3.30), H2/H5, H2/H3, H3/H5, H4/H6, H4/NCH_3 (δ3.30) and H5/NCH_3 (δ 3.30). These data provided enough information to construct a profile for the N-methylleucine spin system.

The HSMQC experiment indicated that H2 was connected to a methine at δ 63.0. The HMBC spectrum showed that the N-methyl protons at δ 3.30 correlated with this methine carbon. These correlations not only assigned H2 and the α-carbon, but also assigned the N-methyl group of the second N-methylleucine residue.

Because there are at least five protons appearing in the region between δ 2.01 and 2.05 region in ^1H NMR spectrum, the HMBC correlations

related to these protons were very complex. H3 showed a strong correlation to the α-carbon at δ 63.0, which confirmed that H3 was part of the second N-methylleucine. The correlation from H4 to a carbon at δ 38.3, rather than δ 38.0, helped assign the former to N-methylleucine (2), and the latter to N-methylleucine (1). The correlation from both H5 and H6 to the same methylene carbon proved the location of this methylene group. The HMBC correlation from H3 to a carbonyl carbon at δ 172.3, which was also correlated to the β-proton of threonine, was recognised. Therefore, the carbonyl carbon was assigned to N-methylleucine (2). The detailed assignment of the HMBC and ROESY data are given in Fig. 4.5.14.

4.5.5 Assembly of the subunits

The 14 units were joined together based on a comprehensive analysis of the data from the HMBC and ROESY experiments. The HMBC analysis gave the linkages for all the above residues. The structure was then confirmed by the ROESY correlations. Because both H2 of valine (1) and the N-methyl proton of N-methylleucine (1) showed correlations to the carbonyl carbon of valine (1) in the HMBC spectrum, the second amino acid residue from the N-terminus was determined to be N-methylleucine (1). H2 of this residue correlated to its own carbonyl carbon at δ 172.8, which in turn was correlated to H2 of the threonine residue. These connections established the structure of the side chain of the entire peptolide, since threonine was the juncture point for the lactone ring. From this point, the linkage goes in both directions.

H2 of threonine was also correlated to another carbonyl carbon at δ 169.2, the threonine CO. The N-methyl protons of the N-Me-β-Ala crossed with the β-C and with the threonine CO. This observation

allowed the extension of the chain to N-Me-β-Ala in one direction. Similarly, the NH of valine (2) correlated with the carbonyl carbon of N-Me-β-Ala and its own α-carbon. An N-methyl proton at δ 3.24 showed correlations to the CO of Val2 and the α-C of MeIle1. In addition, correlations from H2 of MeIle1 to CO of both Val2 and MeIle1 were recognised. The next residue assigned was the β-Ala, by the correlation of the NH with the CO of MeIle1. Correlations from both the NH and H2 of Ile to the CO of β-Ala confirmed that they were adjacent residues. In contrast, the linkage between Ile and MeIle2 was constructed via the correlations from H2 and NCH_3 of MeIle2 to CO of Ile. The H2 of N-MeAla showed cross peaks to two carbonyl carbons, which had been assigned to N-MeAla and MeIle2 respectively. Correlations from the NH of Ala to the CO of MeAla and C2 of Ala put Ala next to N-methylalanine. The 12th amino acid residue was found to be Leu because the NH of its peptide bond showed correlations to the CO of Ala. The NCH_3 of the N-methylleucine (2) correlated with the CO of Leu, which connected these two residues. Finally, observation of a three bond correlation from H3 of threonine to the CO of N-methylleucine 2, the C-terminal amino acid, closed the 36 membered ring. All HMBC correlations are shown in Fig. 4.5.15.

Supporting evidence was also derived from the ROESY correlations as follows H2 (Val1)/NCH_3 (MeLeu1), NH (Thr)/H2 (MeLeu1), H2 (Thr)/NCH_3 (Me-β-Ala), NH (Val2)/H2 (Me-β-Ala), NCH_3 (MeIle1)/H2 (Val2), NH (β-Ala)/H2 (MeIle1), NH (Ile)/H2 (β-Ala), NCH_3 (MeIle2)/H2 and H3 (Ile), NCH_3/H2 (MeIle2), NCH_3 (MeAla)/H2 (MeIle2), NCH_3/H2 (MeAla), NH (Ala)/NH (Leu), NH/H2 and H3 (Leu), NCH_3 (MeLeu2)/H2 (Leu) and NCH_3/H2 (MeLeu2). Complete details of all the ROESY correlations were shown in Fig. 4.5.16.

The complete assignment of the NMR data assembled for theonellapeptolide IIIe has been summarised in Table 4.1.

4.5.6 Conclusion

The work to this juncture has described the entire planer structure of theonellapeptolide IIIe. It was now necessary to determine the relative and absolute stereochemistry of IIIe for both chemical and biological reasons. This problem was solved by X-ray crystallography, in conjunction with a chiral HPLC method.

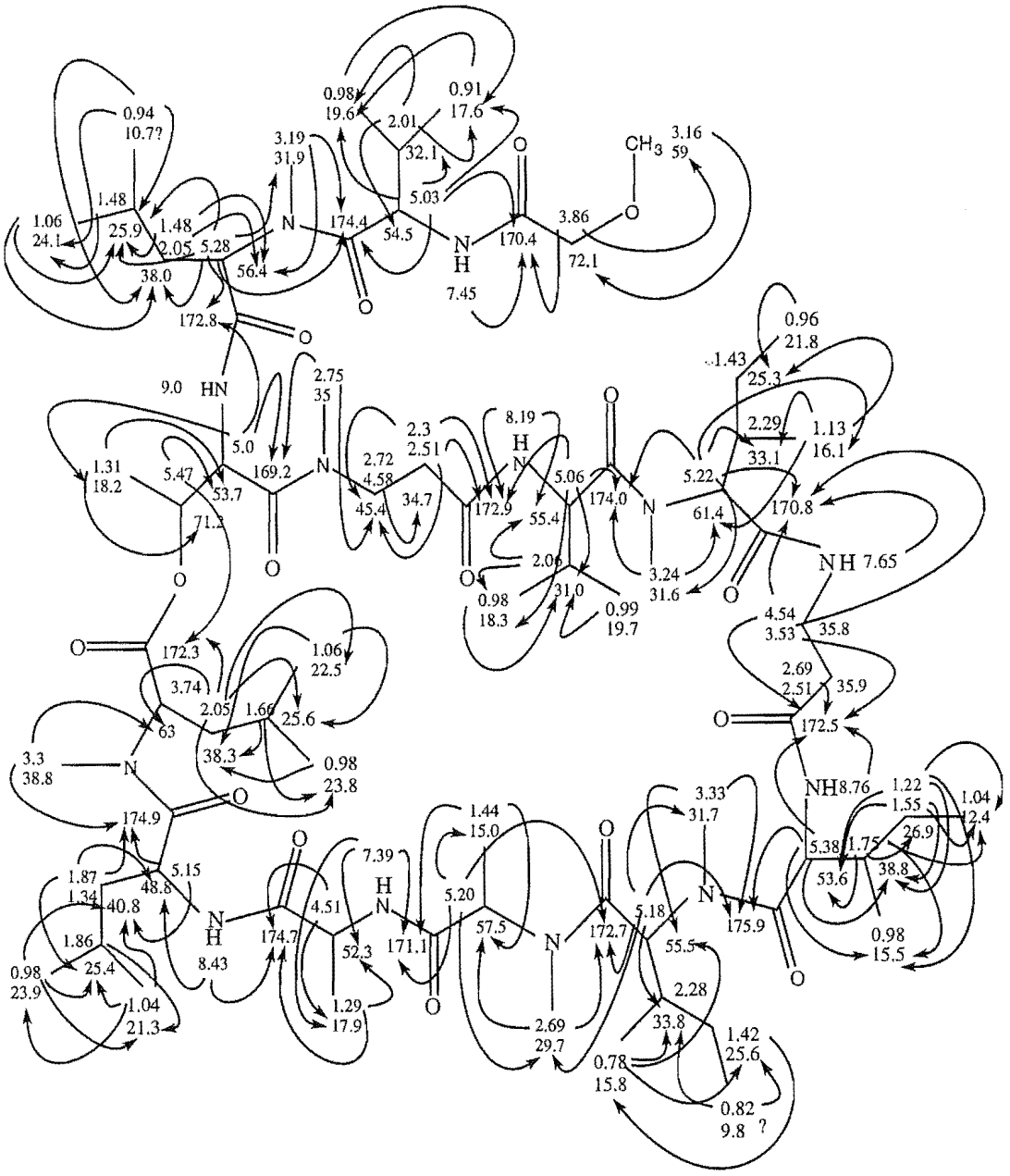


Fig. 4.5.15 HMBC correlations of theonellapeptolide IIIe (**50**)

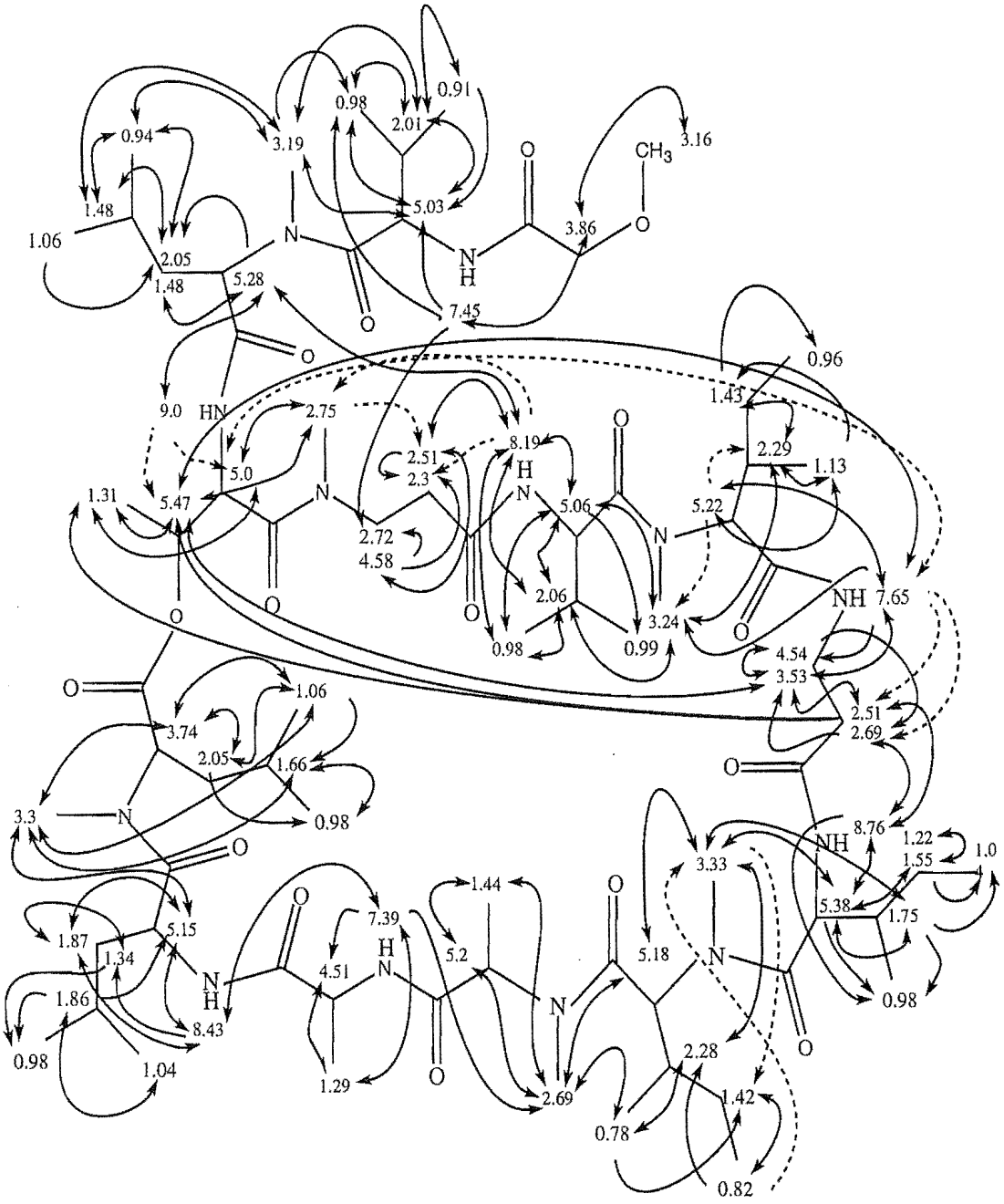


Fig. 4.5.16 ROESY correlations of theonellapeptolide IIIe (**50**)

Table 4.1. Correlated ^{1}H and ^{13}C Spectral Data for Theonellapeptolide IIIe (**50**)

Amino acid (residue #)		δ^a ^{13}C(# H)^b	^{1}H (mult,J(Hz))^c	HMBC^d	ROESY^e
L-Val (1)	1	174.4 (0)			
	2	54.5 (1)	5.03 (m)	174.4, 170.4, 32.1, 19.6, 17.6	3.19, 2.01, 0.98
	3	32.1 (1)	2.01 (m)	54.5, 19.6, 17.6	5.03, 3.19, 0.98, 0.91
	4	17.6 (3)	0.91	19.6	5.03, 2.01
	5	19.6 (3)	0.98	17.6	5.03, 2.01,
	NH		7.45 (d, 9.3)	170.4	5.03, 3.86, 2.72, 0.98
D-*N*-MeLeu (2)	1	172.8 (0)			
	2	56.4 (1)	5.28 (m)	174.4, 172.8, 38.0, 31.9, 25.9	9.0, 8.19, 2.05, 1.48
	3	38.0 (2)	2.05 (m)	56.4, 25.9	1.48, 0.94
			1.48 (m)	25.9, 56.4	5.28
	4	25.9 (1)	1.48 (m)	24.1	3.19, 2.05, 0.94
	5	10.7 (3)	0.94	38.0, 25.9, 24.1	5.28, 3.19, 2.05, 1.48
	6	21.4 (3)	1.06	38.0, 25.9	2.05
	NCH_3	31.9 (3)	3.19 (s)	174.4, 56.4	5.03, 2.01, 1.48, 0.98, 0.94
L-Thr (3)	1	169.2 (0)			
	2	53.7 (1)	5.00 (m)	172.8, 169.2, 18.2	7.65, 2.75, 1.31
	3	71.2 (1)	5.47 (dd, 8.7, 6.3)	172.3, 53.7	7.65, 3.53, 2.75, 1.31
	4	18.2 (3)	1.31 (d, 6.3)	71.2, 53.7	5.47, 5.00
	NH		9.00 (d, 6)		5.47, 5.28, 5.00
N-MeβAla (4)	1	172.9 (0)			
	2	34.7 (2)	2.30 (m)	172.9, 45.4	
			2.51 (m)	172.9, 45.4	8.19, 4.58, 2.30
	3	45.4 (2)	2.72 (m)		
			4.58 (m)	34.7	2.72, 2.51, 2.30
	NCH_3	35.0 (3)	2.75 (s)	169.2, 45.4	7.65, 5.47, 5.00, 2.51
D-Val (5)	1	174.0 (0)			
	2	55.4 (1)	5.06 (m)	172.9, 174.0, 31.0, 18.3	8.19, 3.24, 2.06, 0.99, 0.98
	3	31.0 (1)	2.06 (m)	55.4, 18.3	8.19, 5.06, 3.24, 0.98
	4	18.3 (3)	0.98	31.0	8.19, 5.06, 2.06
	5	19.7 (3)	0.99	31.0	
	NH		8.19 (d, 8.7)	172.9, 55.4	5.28, 5.06, 2.51, 2.06, 0.98
L-*N*-MeIle (6)	1	170.8 (0)			
	2	61.4 (1)	5.22 (m)	174.0, 170.8, 33.1, 31.6, 16.1	7.65, 3.24, 2.29, 1.13
	3	33.1 (1)	2.29 (m)		3.24, 1.43, 1.13
	4	16.1 (3)	1.13 (d, 6.3)	61.4, 33.1, 25.3	5.22, 2.29, 1.43
	5	25.3 (2)	1.43 (m)		2.29, 0.96
	6	21.8 (3)	0.96	25.3	
	NCH_3	31.6 (3)	3.24 (s)	174.0, 61.4	5.06, 2.29, 2.06
β-Ala (7)	1	172.5 (0)			
	2	35.9 (2)	2.51 (m)	172.5	8.76, 5.47, 3.53, 1.31
			2.69 (m)		8.76, 3.53
	3	35.8 (2)	3.53 (m)	172.5, 170.8	7.65, 5.47, 4.54, 2.51
			4.54 (m)	172.5, 170.8	3.53, 2.69
	NH		7.65 (d, 5.4)	170.8	5.47, 5.22, 5.00, 4.54, 3.53, 3.24
D-*allo*-Ile (8)	1	175.9 (0)			
	2	53.6 (1)	5.38 (m)	175.9, 172.5, 38.8, 15.5	8.76, 3.33, 1.75, 1.55, 0.98
	3	38.8 (1)	1.75 (m)	53.6, 26.9, 15.5, 12.4	5.38, 3.33, 1.04, 0.98
	4	15.5 (3)	0.98		
	5	26.9 (2)	1.22 (m)	53.6, 38.8, 15.5, 12.4	1.55
			1.55 (m)	53.6, 38.8, 12.4	5.58, 1.22, 1.04
	6	12.4 (3)	1.04		
	NH		8.76 (d, 9.3)	172.5	5.38, 2.69, 2.51, 0.98
L-*N*-MeIle (9)	1	172.7 (0)			
	2	55.5 (1)	5.18 (m)	175.9, 172.7, 33.8, 31.7, 29.7	3.33, 2.69
	3	33.8 (1)	2.28 (m)		3.33, 0.78
	4	15.8 (3)	0.78 (d, 12)	55.5, 33.8, 25.6	2.69, 2.28, 1.42,
	5	25.6 (2)	1.42 (m)	15.8	0.82
	6	9.8 (3)	0.82 (dd, 9.6, 6.3)	33.8, 25.6	2.28, 1.42
	NCH_3	31.7 (3)	3.33 (s)	175.9, 55.5	5.38, 5.18, 2.28, 1.75, 1.42
L-*N*-MeAla (10)	1	171.1 (0)			
	2	57.5 (1)	5.20 (m)	172.7, 171.7, 29.7, 15.0	2.69, 1.44
	3	15.0 (3)	1.44 (d, 6.9)	171.1, 57.5	5.20, 2.69
	NCH_3	29.7 (3)	2.69 (s)	172.7, 57.5	5.20, 5.18, 1.44, 0.78
L-Ala (11)	1	174.7 (0)			
	2	52.3 (1)	4.51 (m)	174.7, 17.9	1.29
	3	17.9 (3)	1.29 (d, 7.5)	174.7, 52.3	7.39, 4.51
	NH		7.39 (d, 6.3)	171.1, 52.3, 17.9	8.43, 4.51, 2.69, 1.29
D-Leu (12)	1	174.9 (0)			
	2	48.8 (1)	5.15 (m)	174.9, 40.8	8.43, 3.30, 1.87
	3	40.8 (2)	1.34 (m)		1.87, 0.98
			1.87 (m)	174.9, 48.8, 25.4	8.43, 5.15, 1.34
	4	25.4 (1)	1.87 (m)	21.3	1.04, 0.98
	5	21.3 (3)	1.04	40.8, 25.4, 23.9	1.87
	6	23.9 (3)	0.98	40.8, 25.4, 21.3	
	NH		8.43 (d, 8.4)	174.7, 48.8	7.39, 5.15, 1.87, 1.34
D-*N*-MeLeu (13)	1	172.3 (0)			
	2	63.0(1)	3.74 (m)		3.30, 2.05, 1.06
	3	38.3 (2)	2.05 (m)	172.3, 63.0, 25.6, 23.8, 22.5	3.74, 1.06, 0.98
	4	25.6 (1)	1.66 (m)	38.3, 23.8	3.30
	5	23.8 (3)	0.98	38.3	
	6	22.5 (3)	1.06	38.3, 25.6	3.74, 2.05, 1.66
	NCH_3	38.8 (3)	3.30 (s)	174.9, 63.0	5.15, 3.74, 1.66
$MeOCH_2CO$	1	170.4 (0)			
	2	72.1 (2)	3.86 (s)	170.4, 59.0	7.45, 3.16
	OCH_3	59.0 (4)	3.16 (s)	72.1	3.86

[a] Referenced to residual solvent C_6D_6 δ_H=7.27, δ_C=128.4. [b] ^{13}C spectra recorded on a Varian XL300 at 75 MHz. Number of attached H determined by DEPT. [c] ^{1}H spectra recorded on a Varian UNITY 300 at 300 MHz. [d] J_{nxh} = 8.3, 5.0 arrayed mixing times. [e] A mixing time of 0.15s was used.

4.6 Stereochemistry

4.6.1 Crystal Structure of Theonellapeptolide IIIe

A colourless, needle-like crystal of theonellapeptolide IIIe **50** was obtained as a polyhydrate by slow evaporation of an acetone-water (1:1) solution. Data reported here were derived from an irregular chunk of dimensions (0.6 x 0.4 x 0.15 mm). Crystal data: formula $C_{71}H_{127}N_{13}O_{16}/(H_2O)_{18}$, FW=1418/(324); orthorhombic, $P2_12_12_1$, a=12.507(3), b=19.573(4), c=41.416(8) Å, V=10139 $Å^3$, Z=4, Rho=1.142.

The data collection nominally covered a hemisphere. Intensities of 18,590 unique reflections were collected on a Siemens SMART area detector system, fitted with a nitrogen low-temperature gas flow device, using M_oK_α (λ=0.71071 Å) X-radiation. These data were processed using the program SAINT which corrects for LP effects and crystal decay for the duration of the experiment.

The structure was solved by direct methods, and refined using all the unique intensity data in the SHELXTL system of programs. The best converged conventional *R*-factor was 0.12 for 12,029 reflections with $F_0>4\sigma$ (F_0). All numerical data discussed here came from programs in the SHELXTL system as well as from the perspective diagram of the molecule (Fig. 4.6.1 and 4.6.2), which defines the atom labelling scheme used throughout this work.

In addition to the theonellapeptolide molecule shown, there were at least 21 sites for the oxygen atoms of water molecules, 14 of which were at distances appropriate for hydrogen bonding to electronegative O and N

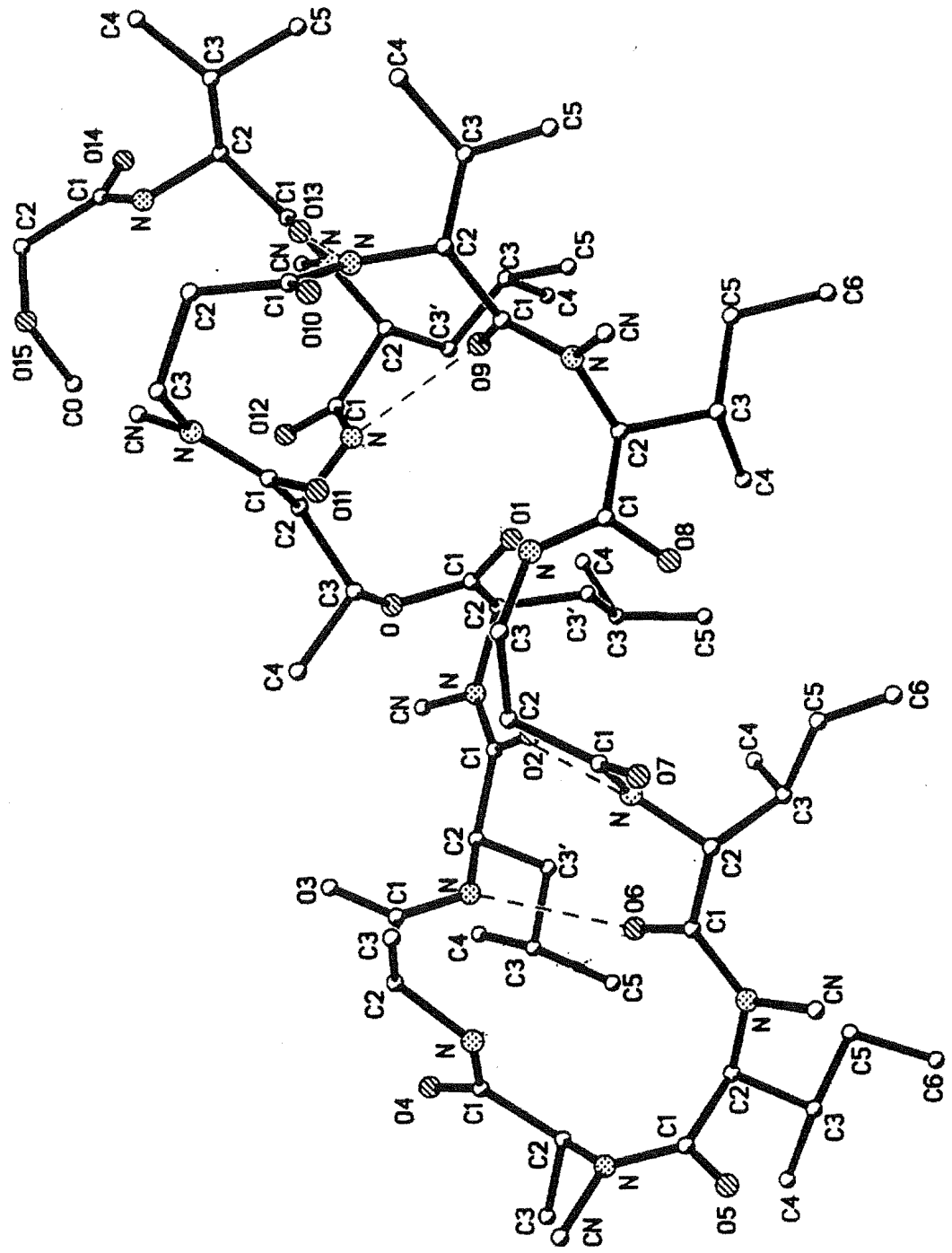


Fig. 4.6.1 Perspective view of the crystal structure of theonellapeptolide IIIe (**50**)

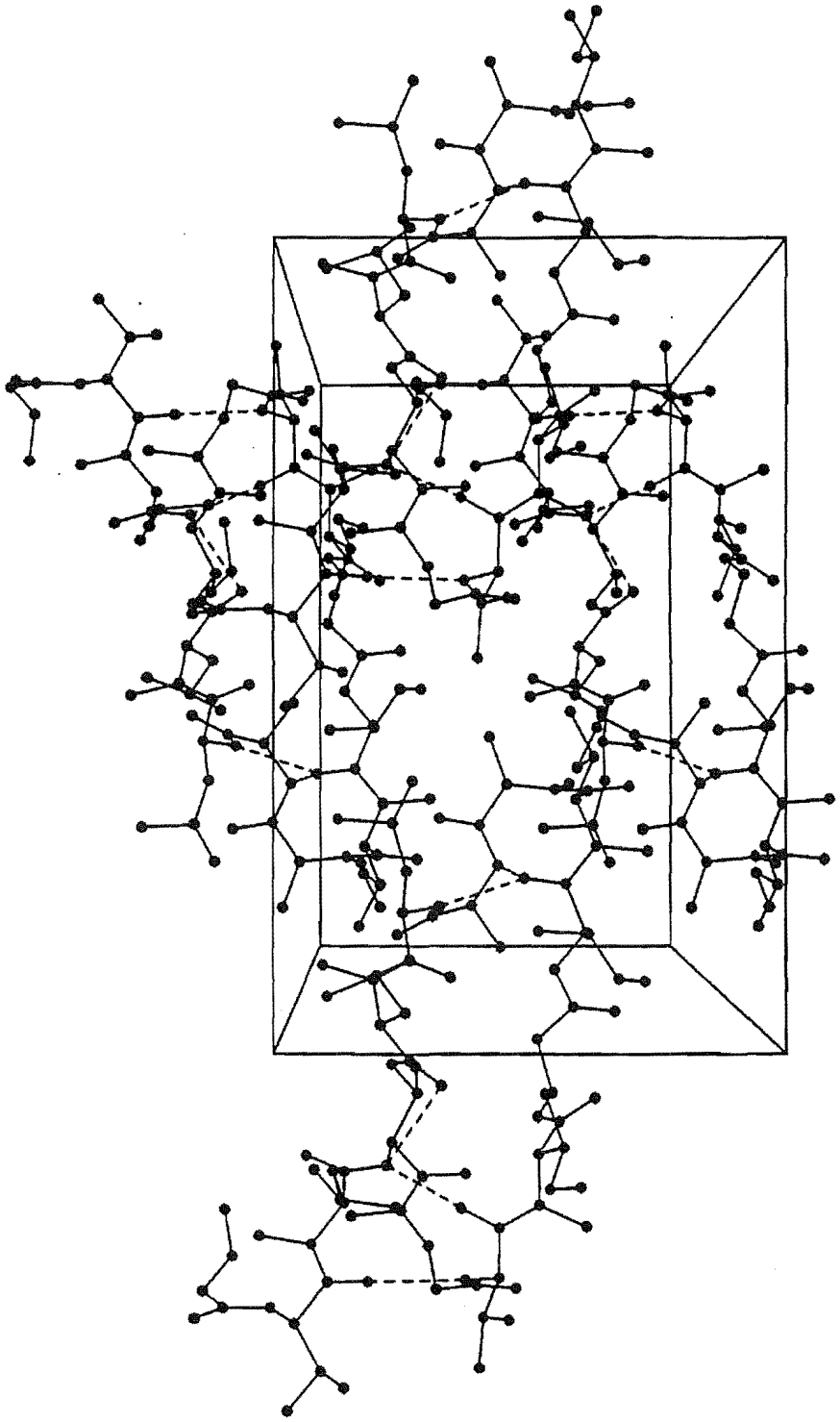


Fig. 4.6.2 Packing structure of theonellapeptolide IIIe (**50**)

atoms of the main molecule. The other seven participated in water-to-water hydrogen bonding. These atomic displacement parameters were consistent with some disorder amongst these water molecules. Six sites, involved in inter-atomic distance interactions between 1.6 and 2.0 Å, must be only partially occupied. These features hindered attainment of better conventional agreement factors. Nevertheless, the molecule of interest was fully characterised by this study.

Figure 4.6.1 shows the crystal structure of **50** with intramolecular hydrogen bonds, as illustrated by the dotted lines. The cyclic tridecapeptide backbone of **50** should be considered to be quite flexible and the configuration of the cyclic part complex, because of the presence of D-amino acids and β-alanine.[154]

The X-ray coordinates were transferred to Chem3D to allow easier viewing of the conformation. The cyclic part of the peptide folded into what is most easily described as a taco shell-like shape (Fig. 4.6.3). In this depiction, five carbonyl groups point to the internal cavity of the taco shell; all the N-methyl groups point down to the base of the taco shell, and the side chains of isoleucine, N-methylisoleucine and leucine protrude over the upper lips. The presence of a β-Ala and an N-MeβAla allow for considerable flexibility in the conformation of the cyclic peptolide ring.

The shape of the molecule is best seen from the stereo diagram (Fig. 4.6.4).

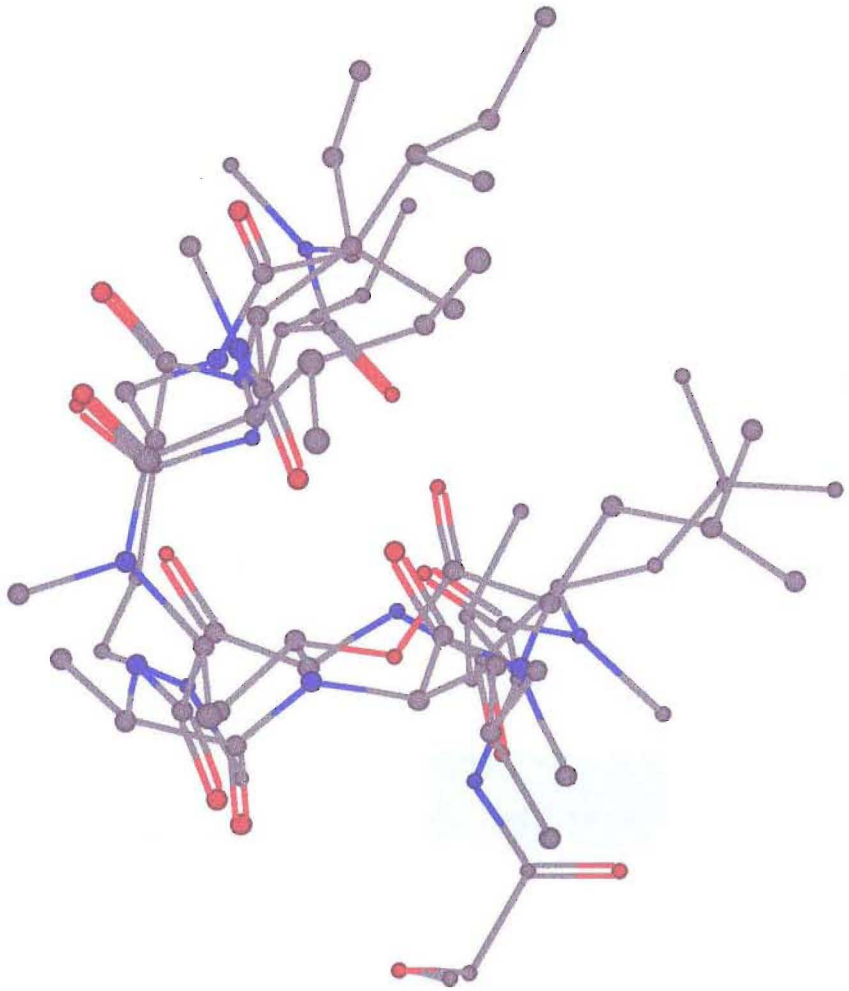


Fig. 4.6.3 Taco shell-like structure from side view of Theonellapeptolide IIIe (**50**)

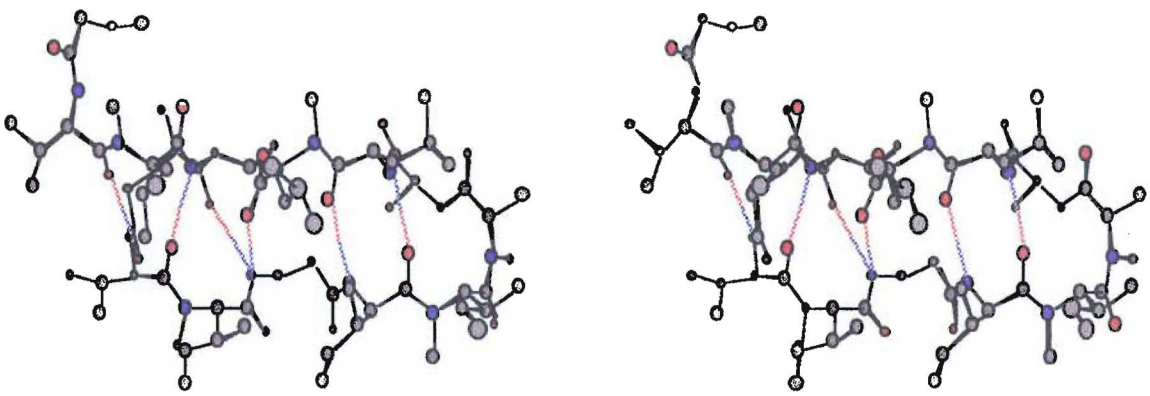


Fig. 4.6.4 Stereoprojection diagram of Theonellapeptolide IIIe (**50**)

4.6.2 Hydrogen Bonds and Secondary Structures

In Figure 4.6.1, all of the carbonyl oxygens present in theonellapeptolide IIIe are involved in hydrogen bonding, either through intramolecular or intermolecular interactions. Oxygens of the carbonyl groups, including O2, O6, O9 and O13 (corresponding to O12, O8, O5 and O1 in Fig. 4.6.5), are directed toward the centre of the macrocyclic peptide ring and are hydrogen bonded to the N6, N2, N11 and N9 (corresponding to N8, N12, N3 and N5 in Fig. 4.6.5) amide hydrogens respectively. The crystal was highly hydrated, and collapsed quickly when it was separated from the solvent. Most of the carbonyl oxygens are directed away from the peptide ring, and are involved in intermolecular hydrogen bonding with the hydrogen atoms of water. For example, O3 has hydrogen bonding with two moles, O4 with one mole , O5 with at least two moles, O7 with three moles, O8 with one mole, O10 with one mole, O12 with one mole, and O15 (these oxygens corresponded to O11, O10, O9, O7, O6, O4, O2 and the methoxy group in Fig. 4.6.5) with one mole of water. In addition, N3, N4 and N13 (corresponding to N11, N10 and N1 in Fig. 4.6.5) were also involved in intermolecular hydrogen bonding with water molecules. It was interesting to note that N3 and N4 associated with the same water molecule, which was also close to the carbonyl oxygen of MeIle2 (O5). This implied that these three functional groups are reasonably close in space (see Table 1.5 in Appendix one). Apparently the conformation was stabilised by both intramolecular and intermolecular hydrogen bonding.

The conformation of theonellapeptolide IIIe can be considered as an antiparallel array of two strands, in which the turns are secured at one end by a β-Ala (residue 4) and at the other by a *cis*-peptide bond [N-

MeIle (residue 9)/N-MeAla (residue 10)]. The resulting conformation is a triple loop ,[154] which is more clearly depicted in the backbone projection diagram (Fig. 4.6.5). The two major loops were defined by the two H-bonds between the *allo*-Ile and the Leu (residues 8 and 12) which are a part of an antiparallel β-pleated sheet structure. The third loop is defined by weak NH...O associations from the NH of the β-Ala (residue 7) to Thr (residue 3) and MeLeu (residue 13) {N7 - O1 [β-Ala (residue 7)-(Thr (residue 3), 3.28Å], N7 - O11 [β-Ala (residue 7)-(N-MeLeu (residue 13), 3.31Å]}. The portion of the peptide outside the macrocycle is static, being held in place by two H-bonds from Val (residue 5) to Thr (residue 3) and Val (residue 1). The MeIle2 and Leu-MeLeu2 sequences form an antiparallel β-sheet structure, stabilised by two hydrogen bonds. The Val1-MeLeu1-Thr and Val2-MeIle1 sequences form a parallel sheet, also stabilised by two hydrogen bonds. On the other hand, the Thr-MeβAla-Val2 sequence forms a loop, in which Me-β-Ala is located at the "bend". The sequence also has a *cis*-MeAla-MeIle amide bond (C2_4-N4-C1_5-C2_5 torsion angle=14.9°).

The previously reported theonellapeptolides I and II differ mainly in variation among the aliphatic side chains. For example, theonellapeptolide Ie,[138] which has the same molecular formula as **50**, has a similar peptolide backbone, but varies in the fifth, ninth, eleventh and thirteenth amino acid residues from the N-terminus, corresponding to substitutions of D-Val for D-Leu, L-N-MeIle for L-N-MeVal, L-Ala for β-Ala, and D-N-MeLeu for D-N-Me*allo*-Ile respectively. These modifications result in a 36-, rather than 37-membered peptolide ring for **50**. An X-ray structure has been reported in a conference abstract[155] for theonellapeptolide Id. Theonellapeptolide Id differs from Ie only in substitution of a β-Ala for a N-Meβ-Ala. Despite the difference in the

five amino acid substituents and the change in ring size, there is a quite remarkable similarity in the overall conformation and H-bonding patterns between theonellapeptolides IIIe (**50**) and Id, which suggests that the theonellapeptolide family quite probably have similar tertiary structures.

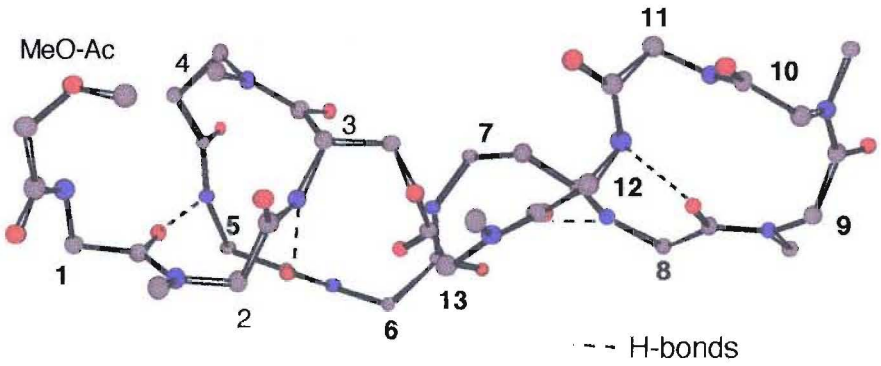


Fig. 4.6.5 Backbone projection diagram of Theonellapeptolide IIIe (**50**)

4.6.3 Absolute Stereochemistry of Theonellapeptolide IIIe

Since it was possible to determine only the relative stereochemistry from the crystal structure, it was necessary to find the absolute configuration of the amino acids by an alternative method. Considering both the availability and the reliability of the method, a chiral HPLC analysis was adopted to achieve this.

The principle of the method is based on the observation that the enantiomers of dansylated amino acids (Dns-Am) can be separated by reversed-phase high-performance liquid chromatography (RPHPLC) when a chiral metal complex was added to the mobile phase. The typical chiral additives include the L-proline-copper complex,[156] the L-valine-copper complex,[157] the L-histidine-copper or L-histidine methyl ester-copper complex,[158] the N,N-di-n-propyl-L-alanine (L-DPA)-copper complex,[159] and the N(p-toluenesulphonyl)-D-phenylglycine-copper complex,[160] as well as the L-2-isopropyl-4-octyl-diethylenetriamine-zinc complex.[161] In all instances, a ternary metal complex was formed between the added chiral metal complex and the enantiomers of the Dns-Am (see Fig. 4.6.6). The difference in the solubility of the L- and D-Dns-Am ternary complexes is the basis for the separation. This difference originates from steric effects in the interaction between the

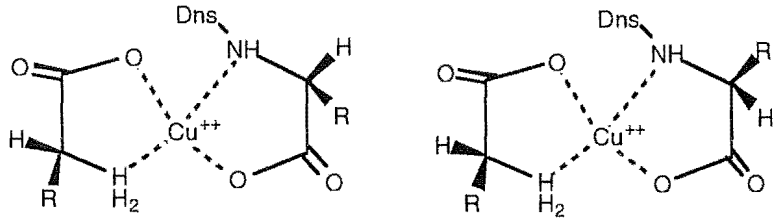


Fig. 4.6.6 Proposed structures of the ternary complexes of L-valine and D,L-isomers of dansylated amino acids with copper (II)

α-substituents of the Dns-Am, and parts of the molecule of the chiral ligand in the metal complex.

In this project, the L-valine-Copper (II) complex was chosen as the chiral additive to the mobile phase. A sample of the peptide was completely hydrolysed with 6M HCl, and the hydrolysate derivatised with Dns-Am using the method of Weinstein and Weiner.[158] The Dns-Am amino acid derivatives were submitted to RPHPLC analysis followed by a co-injection of authentic Dns-L-Thr and Dns-DL-Thr. From these experiments, threonine present in the peptide was found to be S (L). Thus, from the relative stereochemistry obtained from the X-ray data, the configuration of the amino acids present were assigned as: L-Val, D-N-MeLeu(1), L-Thr, D-Val, L-N-MeIle(1), D-*allo*-Ile, L-N-MeIle(2), L-N-MeAla, L-Ala, D-Leu and D-N-MeLeu(2). The structures with all chiral centres at the α-carbon were depicted in Fig. 4.6.7.

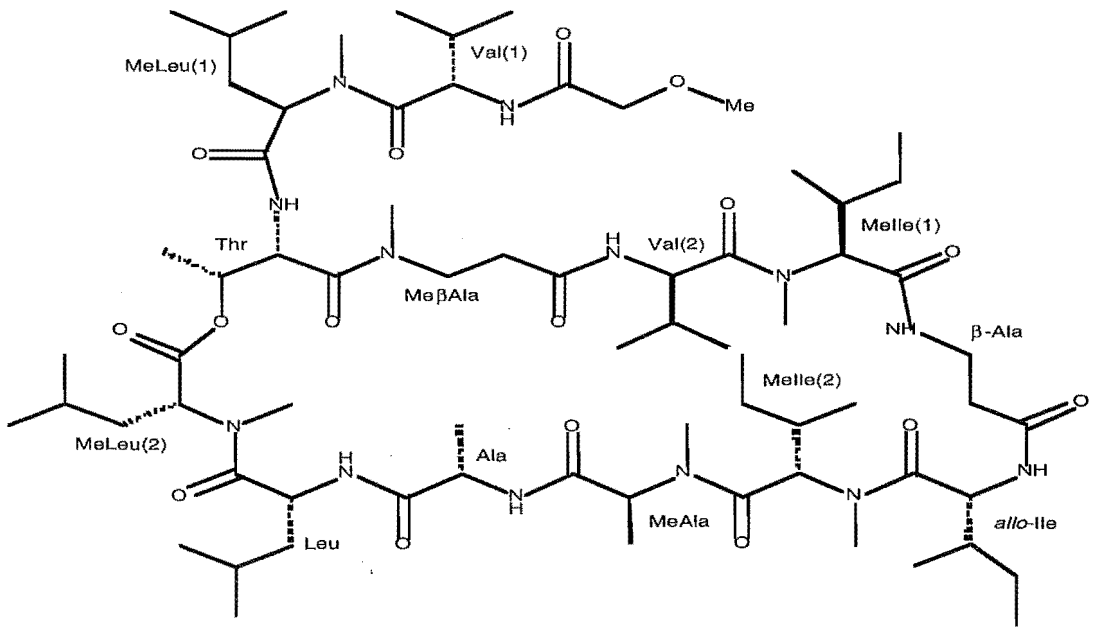


Fig. 4.6.7 Stereochemistry of Theonellapeptolide IIIe (**50**)

4.6.4 Conformational Analysis of Theonellapeptolide IIIe in Solution

Because signal splitting was observed in the ^{1}H NMR spectra in some solvents, e. g. DMSO, and in the ^{13}C NMR spectra for all tested solvents, it was assumed that multiple conformers of this cyclic peptide existed in solution. In contrast, the ^{1}H NMR spectrum did not show the resonances from multi-conformers when using 20% CD_3OD/C_6D_6 as solvent. This implied either a quick interchange between the conformers or that one conformer was dominant in solution.

The solution conformation of theonellapeptolide IIIe was deduced from an analysis of the ROESY data (see Fig. 4.5.16) in conjunction with the X-ray data. Cross peaks of NH (Leu)/NH (Ala) and NH (Ala)/NCH_3 (MeAla) indicated that the N-H bond of Ala was either oriented parallel to the N-H bond of Leu and the N-CH_3 bond of MeAla, or that these three functional groups pointed in same direction. The crystal structure (Fig. 4.6.1) showed that the NH of Ala was in close proximity to both the NH (2.67 Å) of Leu and the NCH_3 (2.98 Å) of MeAla; the dihedral angle for the two N-H and for N-H and N-CH_3 was 30.9° and 91.4° respectively. These data were in agreement with the NMR data. The correlation between the NH of Val2 and MeLeu1 suggested that the ring tail folded to the Val2 part of the ring. The distance between NH (Val2) and H2 (MeLeu1) in the solid state was 3.03 Å which can reasonably explain their relationship. In the ROESY spectrum it is interesting to note the correlations of NH (βAla)/H2 (Thr), NH (βAla)/H3 (Thr), H3 (βAla)/H3 (Thr), H2 (βAla)/H3 (Thr) and H2 (βAla)/H4 (Thr). The Thr and βAla are three residues away from each other. The correlations were attributed to folding of the peptide backbone, which led to these two

amino acid residues becoming closer in space. The proximity of these atoms was also evident in the crystal structure (see Fig. 4.6.1). The measured distances for the five pairs of protons on the βAla/Thr system were 3.59, 1.97, 2.78, 2.33 and 2.63 Å. In addition, the terminal methyl and carbonyl groups of Thr were relatively close to the C2 of MeLeu2 in the crystal structure. This may also account for the abnormally high field chemical shift of the H2 of MeLeu2 in the ^{1}H NMR spectrum. It is possible that the H2 was shielded by these groups, especially from the CO group of Thr.

A correlation from the NH of Val1 to the H3 of MeβAla was observed in the ROESY spectrum. The distance between these two groups (4.85 Å) is too great for efficient interaction. This observation probably arose from the free movement of the side chain in solution.

The evidence from both the X-ray crystal structure and the NMR data suggested that the backbone of theonellapeptolide IIIe adopts a similar conformation in both the solid and solution phases.

This complete structural assignment of theonellapeptolide IIIe provided a good model for studying the structures of the minor peptides isolated from the same sponge. These studies will be described in Chapters 5 and 6.

Chapter Five

Cytotoxic Theonellapeptolide IIIb

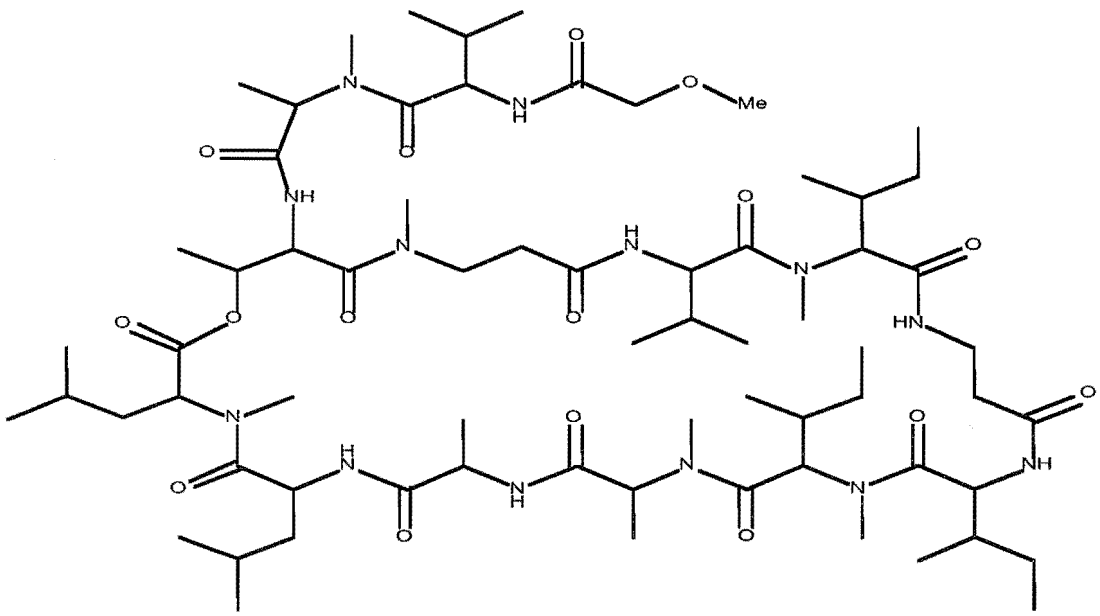
5.1 Introduction

To obtain enough material for the structural study of the minor peptides, a recollection of the sponge, *Lamellomorpha strongylata*, was carried out in April, 1996. A total of 12 kg of fresh sponge was obtained. The sponge (5.96 kg) was extracted by same procedure as described previously. The extract was further separated by bio-assay guided rounds of successive C18 and Diol chromatography. The peptide-containing fractions, located by both bio-assay results and NMR analysis, were combined to give a crude peptide portion.

The peptide fraction was subjected to LH-20 chromatography to separate the peptides from remaining traces of swinholide and calyculin compounds, as well as lower molecular weight impurities. Next, silica gel flash chromatography was used to eliminate the remaining swinholide. Finally, semi-preparative reverse phase HPLC was used to separate the peptide mixture. A total of nine fractions were collected. ^{1}H NMR analysis indicated that seven of them contained peptide components. Fractions 2-6 were purified by repeated RPHPLC to yield the theonellapeptolides IIIa, b, c, d and e respectively. The seventh peptide eluted before IIIb. It was designated as IIIg, due to the failure to recognise it in previous analytical HPLC analysis. In this chapter, the detailed structural elucidation of theonellapeptolide IIIb, will be presented.

5.2 Characterisation of Theonellapeptolide IIIb

Theonellapeptolide IIIb (**52**) was mildly cytotoxic against the P388 cell line (IC_{50} 9.9 μg/mL). The structure was elucidated by the following procedure: an examination of the ^{1}H and ^{13}C NMR spectra to recognise the peptide character; IR spectrum and ninhydrin test to gain insight into the properties of the N- and C-termini; amino acid analysis by GC/MS coupled with FAB MS/MS analysis of the ring-opened peptide for sequencing the peptide chain; TOCSY, DEPT, HSMQC and HMBC spectra to establish individual spin systems and identify the isobaric amino acid residues; detailed analysis of the HMBC data to obtain the linkage of the amino acid fragments and finally the complete assignment of ^{1}H and ^{13}C NMR resonances.



Theonellapeptolide IIIb **52**

Theonellapeptolide IIIb (**52**), obtained as an amorphous powder, $[\alpha]_D$ -49.1° (c=0.745, MeOH), showed an $[M+H]^+$ ion at *m/z* 1377. High-resolution FAB mass spectral data on the $[M+Cs]^+$ yielded the molecular

formula $C_{68}H_{121}O_{16}N_{13}Cs$ (*m/z* 1508.8134, Δ +2.5 mu), indicating 15 double bond equivalents. The difference of 42 amu from theonellapeptolide IIIe corresponds to a difference of C_3H_6. The peptolide nature of **52** was suggested by its ^1H and ^{13}C NMR spectra, showing seven amide NH, 13 amide carbonyl groups and one ester carbonyl group, as shown in Table 5.1. The ^1H NMR spectrum exhibited seven methyl singlets between δ 2.68 and 3.32, corresponding to the resonances of N- or O-methyl groups (see Fig. 5.2.1). The IR spectrum showed typical peptide bands at 3352, 1672, 1626, 1522 cm^{-1} and a lactone band at 1734 cm^{-1}. Amino acid analysis, by GC/MS, identified 13 amino acids. These accounted for 13 degrees of unsaturation, the 14th arising from a methoxyacetate group (observation of a singlet methylene signal at δ 3.84 in the ^1H NMR spectrum and an additional carbonyl group at δ 170.7 in ^{13}C NMR spectrum). The 15th double bond equivalent was the lactone ring. No other elements of unsaturation were detected in the ^1H and ^{13}C NMR spectra. A negative result from the ninhydrin test suggested that the N-terminus of the peptide was blocked. Therefore, a theonellapeptolide skeleton was best established from the initial spectral data comparison with theonellapeptolide IIIe.

Sequencing of the ring-opened peptide (methanolysis product) gave a similar arrangement of amino acid residues in IIIb, compared with IIIe, except for the replacement of an N-methylleucine (1) with N-methylalanine. All seven amino acid spin system corresponding to those with free NH group were identified by tracing down the TOCSY correlation spectra as described earlier. The spin systems of the N-

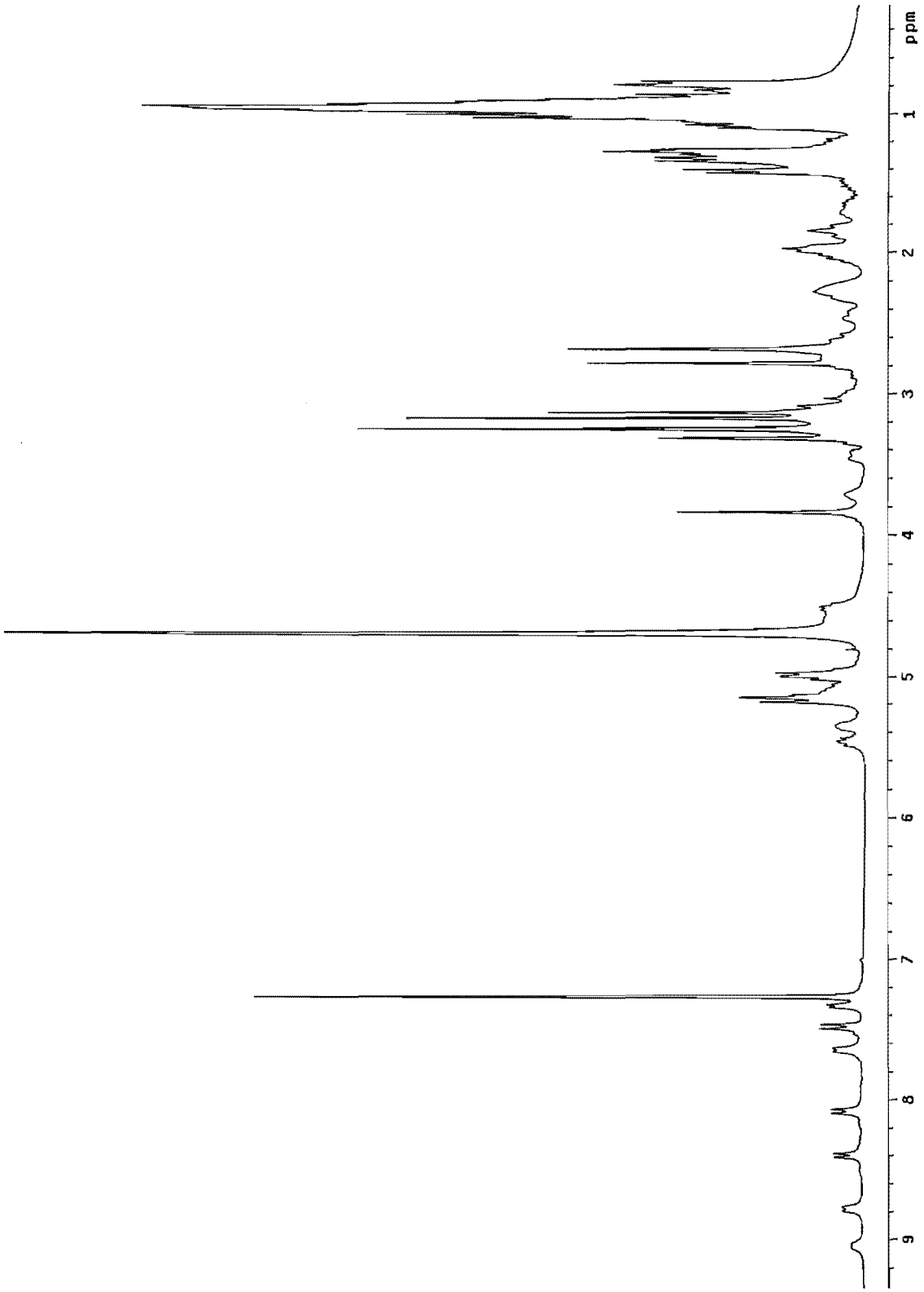


Fig. 5.2.1 ^{1}H NMR spectrum of theonellapeptolide IIIb (**52**)

methyl amino acids, except for N-methylalanine (1), were determined by comparison with the structure of IIIe, and by careful analysis of the TOCSY data for IIIb. Determination of the N-methylalanine (1) residue was not straightforward because the resonances overlapped with other residues, and there was no corresponding residue in IIIe. A comparison between the ^{1}H NMR spectra of IIIb and IIIe led to the recognition of an additional doublet at δ 1.36 in IIIb. This new signal was presumed to be the methyl resonance of N-methylalanine (1). ^{1}H-^{1}H decoupling experiment were used to verify the assumption. When the signal at δ 1.36 was irradiated, the multiplet around δ 5.12 sharpened, when the multiplet at δ 5.12 was irradiated, the doublet at δ 1.36 collapsed into a sharp singlet. These results confirmed the spin system of N-methylalanine (1).

The connectivities of the individual amino acid residues were primarily established by HMBC analysis. The two bond correlations between NH's and carbonyl carbons were readily recognised. Using these as the starting point, the relationships with adjacent N-methyl amino acids were determined by observation of the correlations to the same carbonyl carbon, both from the α-proton and from the N-methyl protons of the adjacent residue. Complete ^{1}H and ^{13}C NMR resonance assignments are summarised in Table 5.1.

5.3 Amino acid analysis

Theonellapeptolide IIIb was completely hydrolysed with 6M HCl, followed by derivatization. The N-TFA-n-butyl derivatives of the hydrolysate were subjected to GC and GC/MS analysis. Several unusual N-methyl amino acids were synthesised for identification purposes (see Experimental). The individual component amino acids were identified by comparison against the retention times of the standard amino acid derivatives and by interpretation of the MS spectra. The detailed GC/MS analysis was similar to that described in Section 4.4. A total of 13 amino acids were identified, including 2 x Val, 2 x N-MeAla, 2 x N-MeIle, 1 x N-MeLeu, 1 x Thr, 1 x Leu, 1 x Ile, 1 x Ala, 1 x β-Ala and 1 x N-Me-β-Ala.

5.4 Sequencing of the Ring-Opened Peptide

Because theonellapeptolide IIIb (**52**) is a depsipeptide, the lactone ring can be opened by methanolysis to form the linear peptide methyl ester **53**. The same hydrolysis procedure, used previously in Section 4.5, was adapted for this peptolide. After separation/purification of the hydrolysate by RPHPLC, the major product was subjected to FAB MS/MS analysis.

LRFABMS showed an $[M+H]^+$ ion at *m/z* 1409, 32 amu higher than the intact peptolide, which corresponded to the methyl ester of the ring-opened peptide. This also confirmed the methanolysis of the lactone, rather than hydrolysis to a free carboxylic acid moiety, which would have afforded an increase in mass of 14 amu.

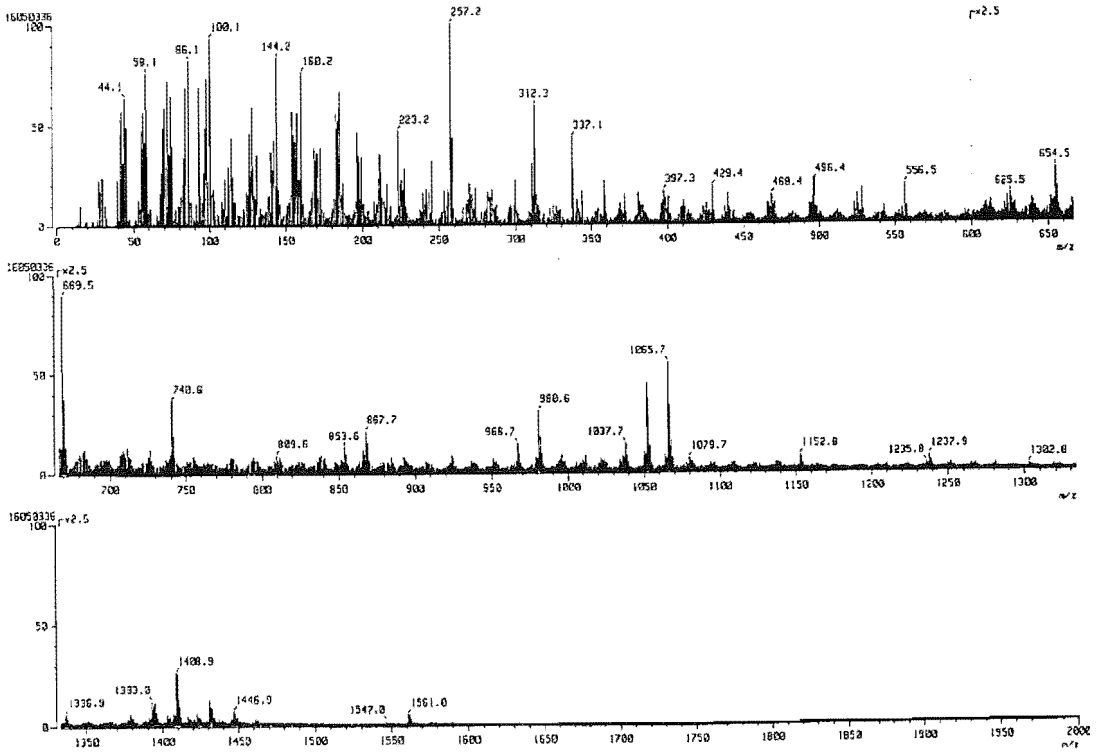
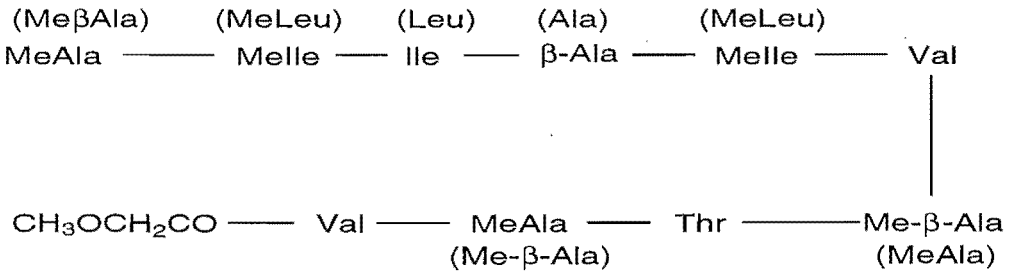
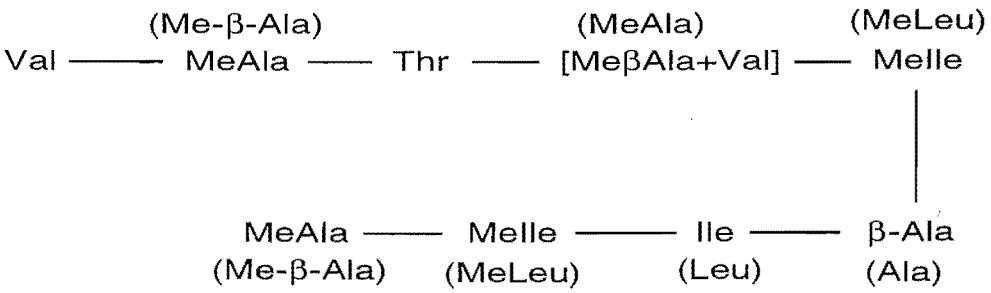


Fig. 5.4.1 FABMS spectrum of compound **53**

The collision activated decomposition (CAD) mass spectrum of compound **53** produced two series of the ions: (a) *m/z* 1066, 981, 854, 741, 670, 543, 358, 257 and 172; (b) 1238, 1153, 1052, 967, 868, 741, 556 and 429 (Fig.5.4.1). On the supposition that these ions had been formed by either B or Y type fragmentation, the peptide fragments **53a** and **53b** were mapped out from these data.



Peptide fragment **53a**



Peptide fragment **53b**

Comparison of fragment **53a** with **53b** indicated that fragment **53b** corresponded to elimination of a methoxyl-acetate valine moiety (171 amu) from **53a**. Fragment **53a** showed a 344 amu difference from the intact peptide **53**, which corresponded to a tripeptide segment. The postulated fragments were further confirmed by MS/MS analysis of the selected parent ions, which provided more connectivity information.

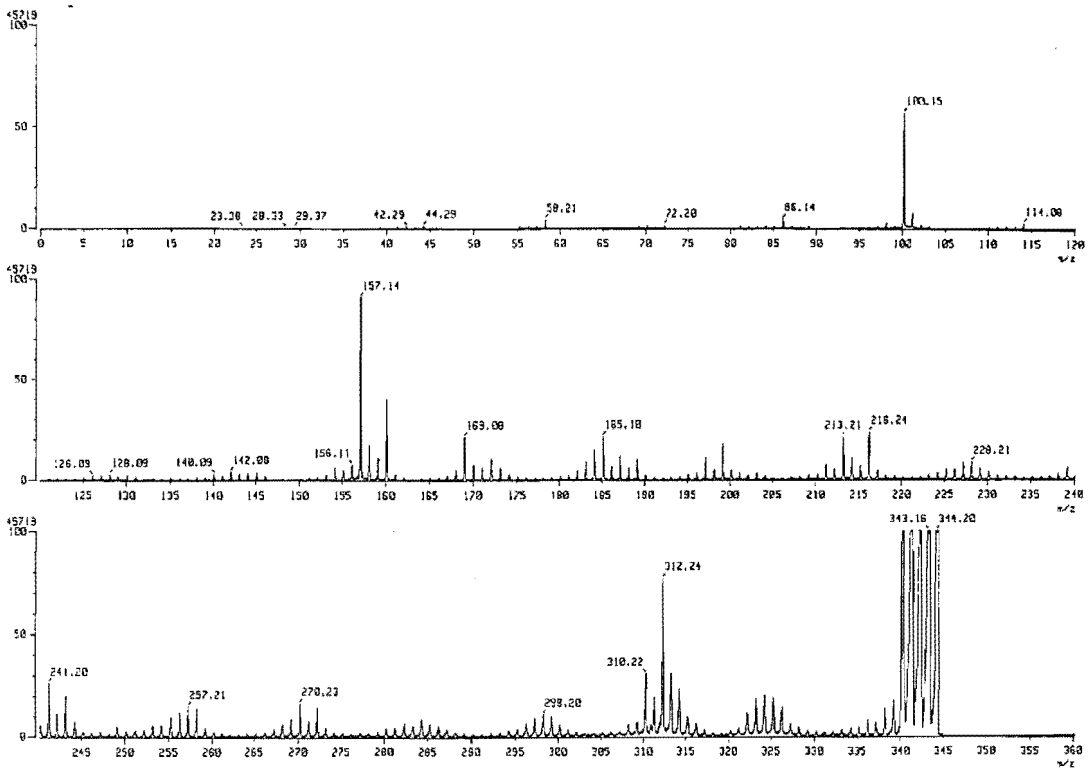


Fig. 5.4.2 Link scan at *m/z* 344

A link scan from *m/z* 344 (Fig. 5.4.2) gave the same daughter ions as those from scanning the same mass in theonellapeptolide IIIe (see Fig. 4.4.4), which suggested that IIIb contains the same tripeptide fragment at its C-terminus, i.e., CH_3O-Leu-Leu-Ala. An intense 344-32 signal (CH_3OH) confirmed that the tripeptide was the C-terminal fragment.

The N-terminal sequence of the linear peptide was confirmed as follows. Peaks at *m/z* 340, 257, 187 and 172 were daughter ions of *m/z* 358 (Fig. 5.4.3). These were formed by the successive loss of H_2O, Thr, CH_3OCH_2CO-Val and Thr-N-MeAla fragments from the parent ion. It was important to note that the *m/z* 257 ion arose from *m/z* 357 by elimination of a threonine residue. When the *m/z* 257 ion was scanned, strong *m/z* 172 and medium intensity of *m/z* 100 signals were observed (Fig. 5.4.4). These ions arose from elimination of a N-methylalanine followed by elimination of the methoxyl acetate from the N-terminus.

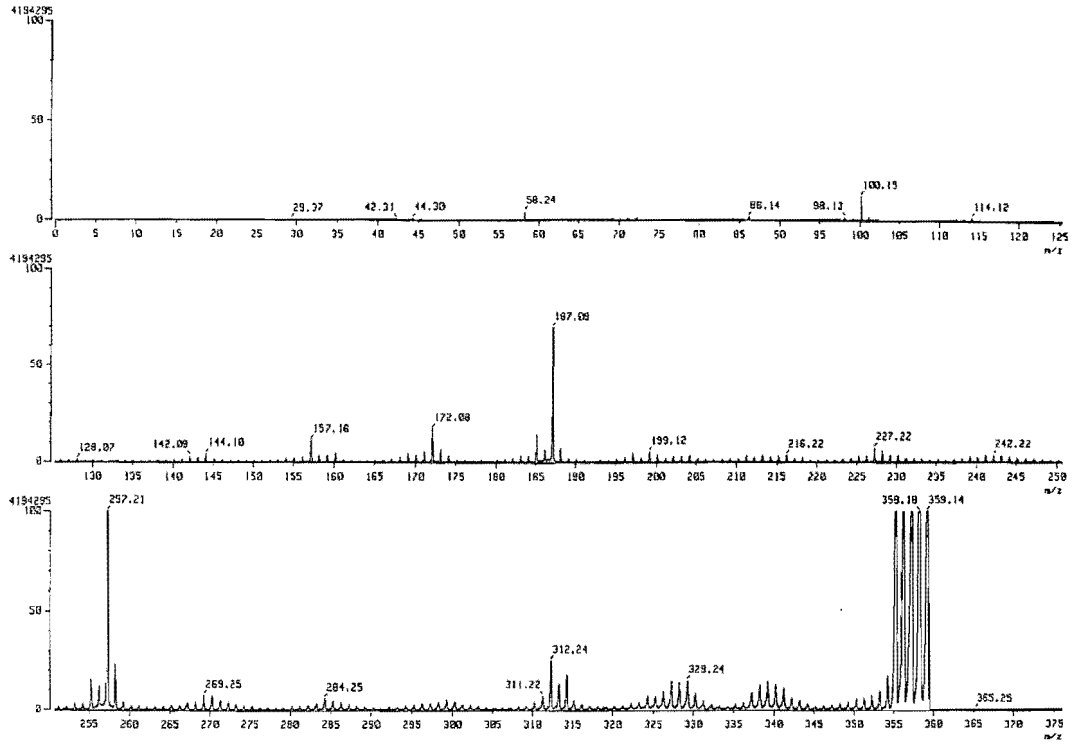


Fig. 5.4.3 Link scan at *m/z* 358

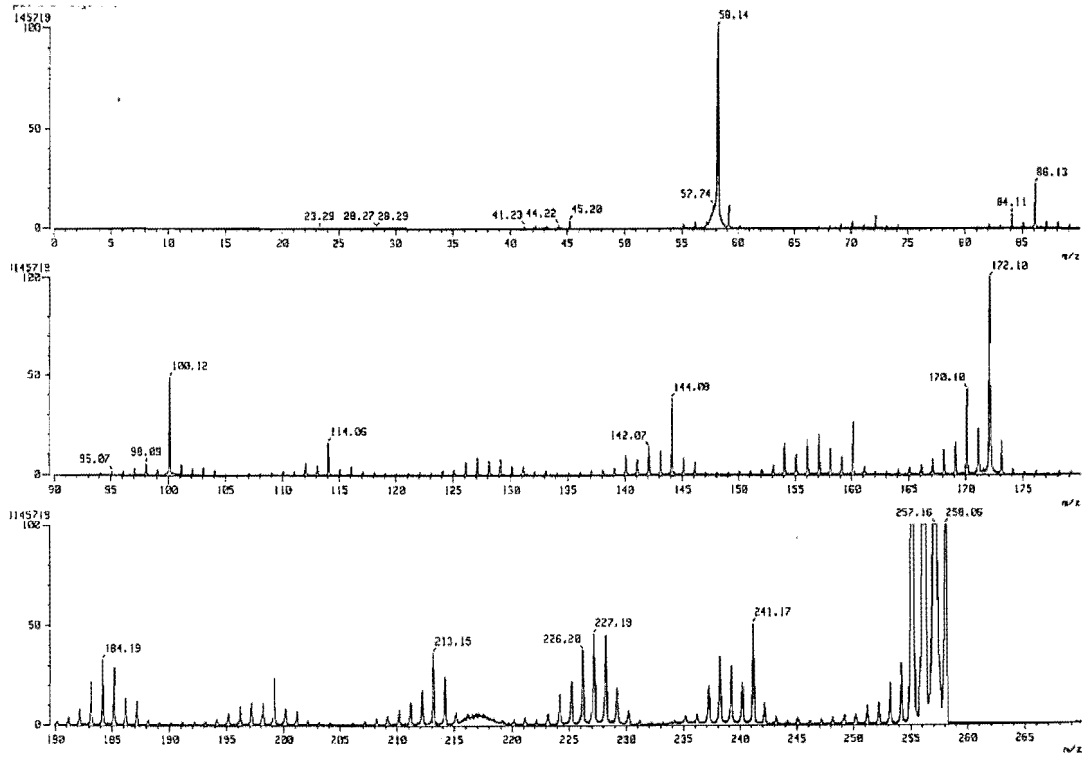


Fig. 5.4.4 Link scan at *m/z* 257

The CAD spectrum of *m/z* 257 also produced immonium ions of N-methylalanine, or N-methyl-β-alanine (*m/z* 58) and valine (*m/z* 72). The presence of these amino acids established the side chain of the cyclic peptolide as CH_3OCH_2CO-Val-N-MeAla (or N-Me-β-Ala).

The daughter ion spectrum from the ion at *m/z* 1237 ion (Fig. 5.4.5) revealed ions at *m/z* 1051, 966, 867, 740, 669 and 556 which were assigned as [MeβAla-Val-MeIle-βAla-Ile-MeIle-Ala-Leu-MeLeuOCH_3]$^+$, [Val-MeIle-βAla-Ile-MeIle-Ala-Leu-MeLeuOCH_3]$^+$, [MeIle-βAla-Ile-MeIle-Ala-Leu-MeLeuOCH_3]$^+$, [βAla-Ile-MeIle-Ala-Leu-MeLeuOCH_3]$^+$, [Ile-MeIle-Ala-Leu-MeLeuOCH_3]$^+$ and [MeIle-Ala-Leu-MeLeuOCH_3]$^+$ fragments respectively. Besides extending the peptide chain from the N-terminus to the C-terminus, this series of ions also suggested that N-methylalanine was adjacent to threonine, β-alanine to N-methylisoleucine (1), and the β-alanine was followed by isoleucine.

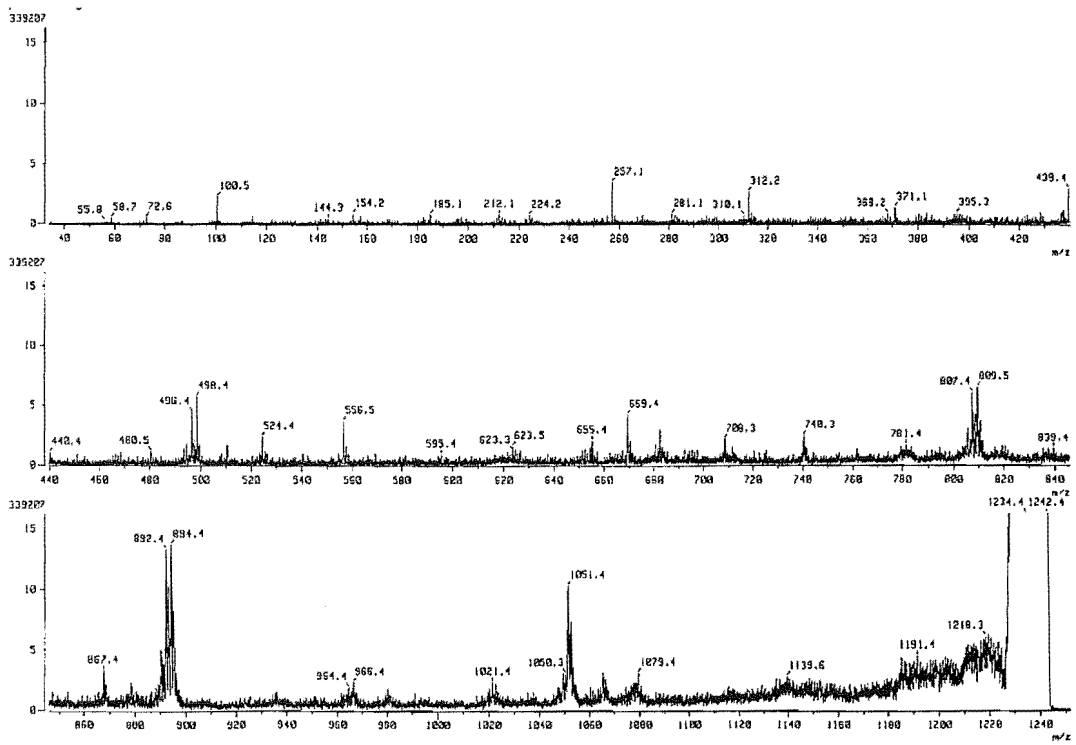


Fig. 5.4.5 Link scan at *m/z* 1237

In this sequence deduced from the FABMS spectrum of compound **53**, two common ions at *m/z* 740 and 669 should be formed by cleavage from either the N- or C-terminus (the Y_i and B_i ion series, respectively). Confirmation of the origin of these two ions would help verify the sequence. Since *m/z* 740 and 669 had been found as the daughter ions from the *m/z* 1237 ion, which had resulted from loss of the N-terminal methoxyacetate and valine, the origin from the C-terminal cleavage can be deduced by link scans at *m/z* 1065, 853 and 669 ions.

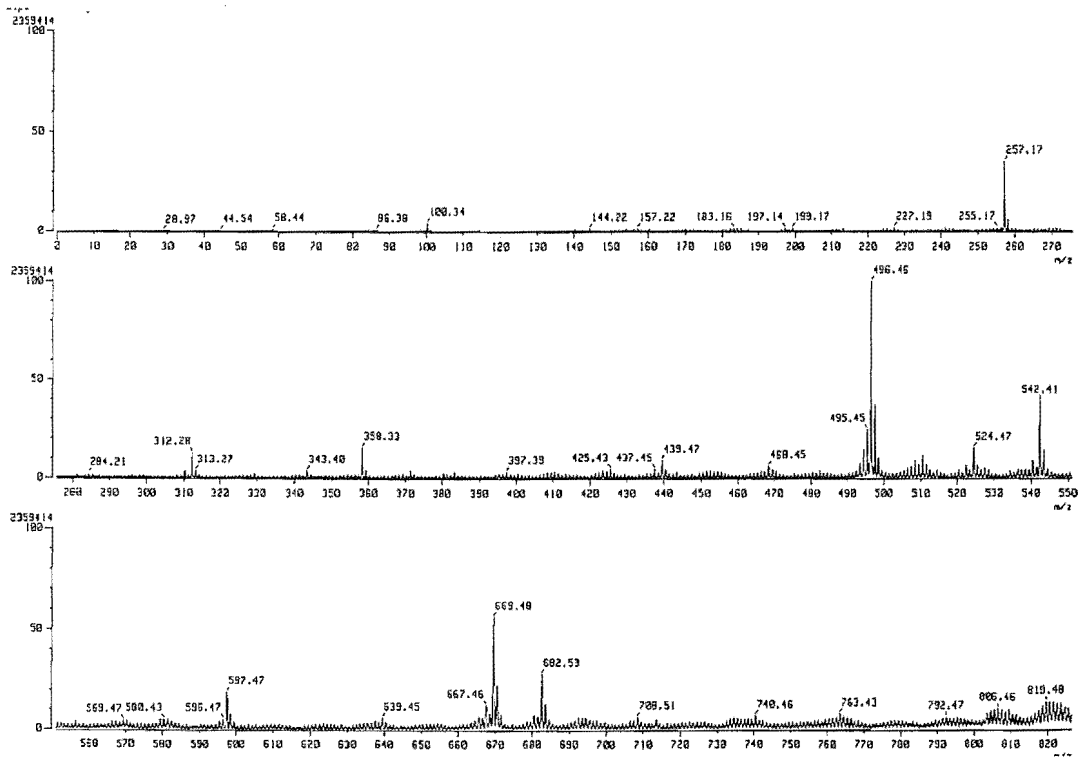


Fig. 5.4.6 Link scan at *m/z* 853

Link scans from the *m/z* 1065 ion gave abundant daughter ion signals at *m/z* 980, 669 and 257, as well as a reasonably intense *m/z* 358 ion, which can be attributed to B_9, B_6, B_2 and B_3 ions. However, the intensities of some expected daughter ions at *m/z* 853, 740, 542 and 443 were too low to afford sufficient evidence for the proposed sequence. Fortunately, ions at *m/z* 740 and 542 were picked up from the link spectrum at *m/z* 853

(Fig. 5.4.6). Although the *m/z* 740 ion was weak, the appearance was that of a true signal. In addition, scanning the ion at *m/z* 669 yielded *m/z* 542, 443, 358 and 257 ions, which complemented the suggested sequence. Thus, sequencing the linear peptide from both the N- and C-termini was accomplished with the same result. The important Bi and Yi ions and the intact sequence of compound **53** are depicted in Fig. 5.4.9.

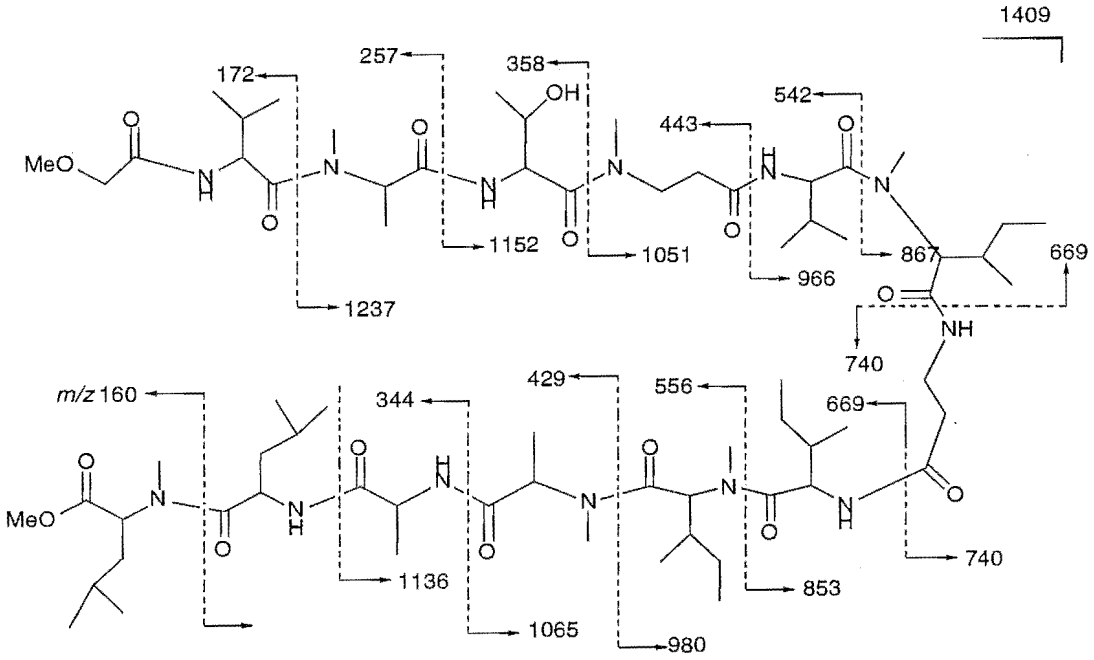
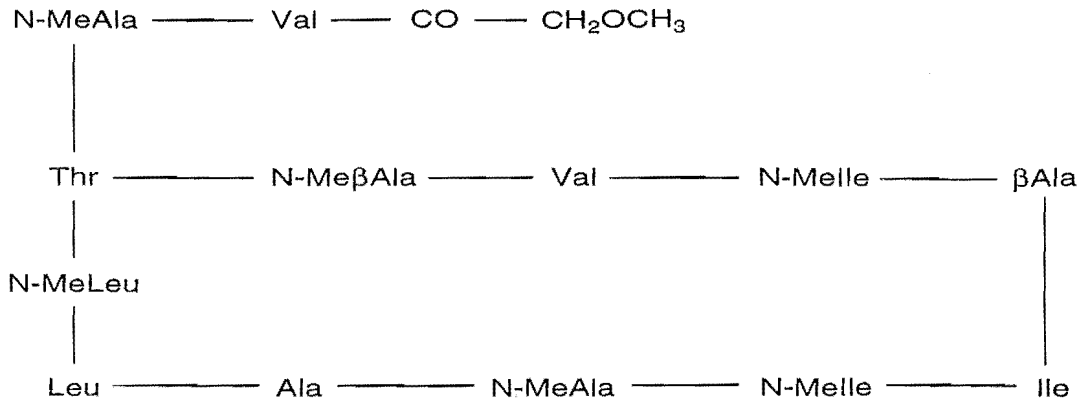


Fig. 5.4.9 FAB MS/MS fragmentation of compound **53**

Because only one hydroxy amino acid was recognised from the GC/MS analysis and by NMR spectra, the lactone ring must be closed by an ester linkage between the carboxyl group of the C-terminal residue and the hydroxyl group of threonine. So, the structure of the intact peptolide can be deduced as **52** (Fig. 5.4.10). However, all the amino acids that are isobaric are exchangeable. Detailed 1D and 2D NMR analysis was now necessary to resolve these ambiguities.



Intact sequence of theonellapeptolide IIIb (**52**)

N.B. Alanine and β-Alanine, N-methylalanine and N-methyl-β-alanine, N-methylleucine, and N-methylisoleucine and N-methyl-*allo*-isoleucine, leucine, isoleucine and *allo*-isoleucine are isobaric isomers and their position are exchangeable.

5.5 Complete Assignment of the Structure of Theonellapeptolide IIIb

The assigned chemical shifts and coupling constants of theonellapeptolide IIIb are given in Table 5.1. Seven amide protons were detected at δ 9.05 (1H, br d, *J*=ca. 6.3 Hz), 8.76 (1H, d, *J*=8.4 Hz), 8.40 (1H, d, *J*=8.4 Hz), 8.07 (1H, d, *J*=8.1 Hz), 7.64 (1H, d, *J*=6.6 Hz), 7.49 (1H, d, *J*=9.3Hz) and 7.32 (1H, d, *J*=5.7 Hz) in the ^{1}H NMR spectrum. Six N-methyl and one *O*-methyl groups were found at δ 3.32. 3.25, 3.18, 3.14, 2.79 and 2.68 as sharp singlets. The signal at δ 3.18 was again assigned to a methoxyl group because of its attachment to a quaternary carbon at δ 59.2 (HSMQC and DEPT analysis). The assignments of the spin systems corresponding to the individual amino acid residues were established from either the analysis of TOCSY spectra, or comparison with the spectral data of theonellapeptolide IIIe, and confirmed by detailed analysis of the DEPT, HSMQC, HMBC and ROESY spectra, as well as ^{1}H-^{1}H decoupling experiments.

The methylene protons of the methoxyl acetate of IIIb appeared as an isolated singlet at δ 3.84 in the ^{1}H NMR spectrum, and showed correlations to both its own carbonyl carbon (δ 170.7) and a methoxyl carbon in the HMBC spectrum. This was virtually identical to the behaviour of IIIe. At the same time, a correlation from the methoxyl protons to the methylene carbon (δ 72.0) were also observed. This assignment provided a firm starting point and allowed the N-terminal residue to be located in turn.

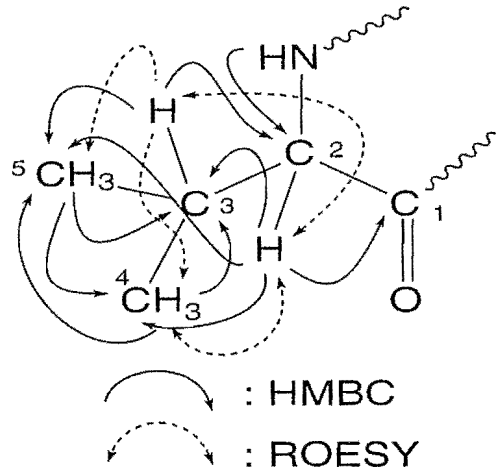


Fig. 5.5.1 Selected HMBC and ROESY correlations of Val1

The partial structure of valine 1 is shown in Fig. 5.5.1. In the TOCSY spectrum, the amide proton of Val1, coupled with signals at δ 4.99, 2.06, 1.01 and 0.94, had been defined as the N-terminal residue by the observation of two bond correlations to the methoxy acetyl carbonyl and to its own α-carbon (δ 54.6). Since the α-proton at δ 4.99 was correlated with two carbonyl carbons at 170.7 and 173.9, the latter signal was assigned as the carbonyl carbon of Val1. The correlations between CH_2 (methoxyl acetate)/NH, CH_2/H2, NH/H2 and H2/H3 in the ROESY spectrum confirmed the moiety.

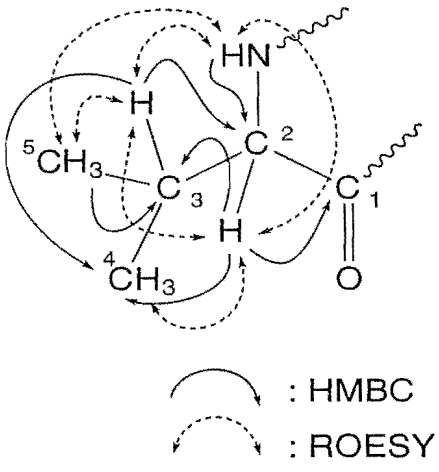


Fig. 5.5.2 Selected HMBC and ROESY correlations of Val2

The second valine residue (Fig. 5.5.2) was recognised by the correlations to δ 4.98, 2.0, 0.91 and 0.98 from the amide proton at δ 8.07 in the TOCSY spectrum. Both the NH and the αH were correlated with the same carbonyl carbon at δ 173.3, but the αH was also correlated to a second carbonyl carbon at δ 174.1. This latter peak was assigned to Val2. The β-carbon was located by observation of the one bond H-C coupling in the HSMQC spectrum. One of the side chain methyl groups was characterised by a two bond correlation from the methyl proton to this β-carbon, and a correlation from the βH to a primary carbon at high field (δ 19.6) provided evidence for the other methyl group of valine. All individual protons were located from the HMBC data coupled with analysis of the HSMQC spectrum, and verified by ROESY correlations.

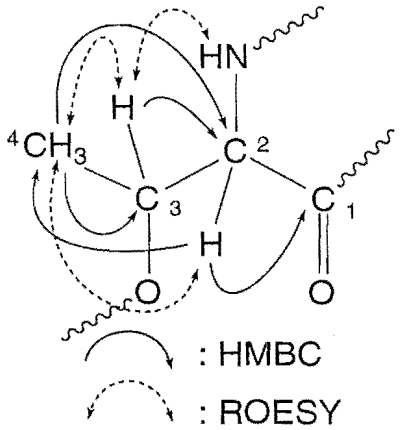


Fig. 5.5.3 Selected HMBC and ROESY correlations of Thr

The next residue examined was threonine. In the TOCSY spectrum, threonine gave a very distinctive trace. Differentiation between the H2 and the H3 was achieved by comparison of two pairs of one bond H-C coupling in the HSMQC spectrum. Since a proton at δ 5.45 was coupled to an oxygen-bearing carbon at δ 70.8, a typical value for the β-carbon of threonine, the other H/C (δ 4.95/53.5) pair can be placed at the α-position. The carbonyl carbon of threonine was located by correlations

both from its own α-proton and from the protons of the adjacent residue (see Fig. 5.5.3).

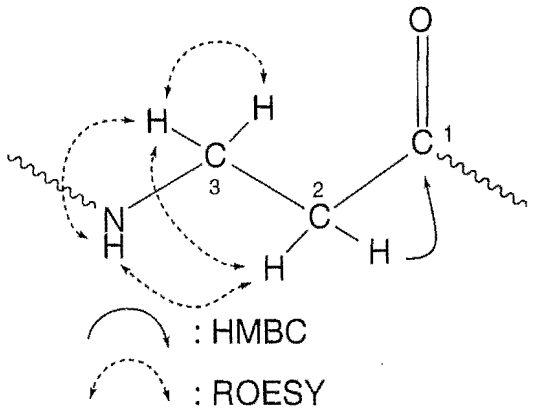


Fig. 5.5.4 Selected HMBC and ROESY correlations of β-Ala

For assignment of β-alanine it was noted in the TOCSY spectrum that the NH at δ 7.64 was coupled with signals at δ 4.55, 3.42, 2.69 and 2.40. These four protons were part of two methylene groups (HSMQC spectrum), which suggested a β-alanine residue. The ROESY cross peaks of H3/H3, H3/NH and H2/H3 established their relative positions within the residue. In the HMBC spectrum, correlations were observed to the carbonyl of β-alanine (δ 172.8). The same carbonyl group was also correlated with the H2 of the neighbouring residue.

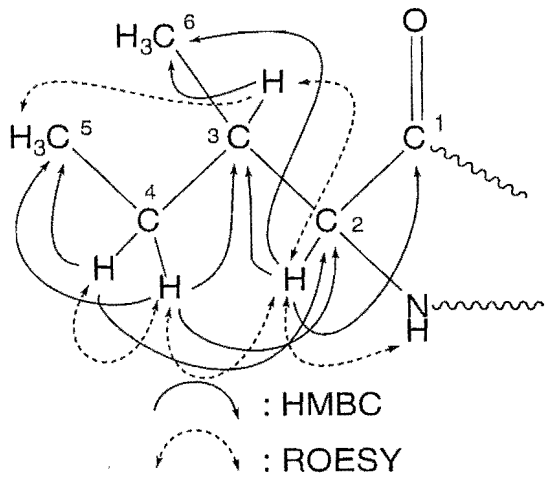


Fig. 5.5.5 Selected HMBC and ROESY correlations of Ile

The deduced partial structure of isoleucine is shown in Fig. 5.5.5. In this case the NH did not show correlations with other protons in the spin system by TOCSY. However, when tracing from the β-proton at δ 1.73, all cross peaks except the NH in TOCSY spectrum were found. In the HMBC spectrum, the α-proton showed correlations to the carbonyl of β-Ala, its own carbonyl (δ 175.9), the β-carbon and a methyl carbon. The last correlation was also good evidence for an isoleucine residue. Two protons in the spin system with different chemical shifts (δ 1.18, 1.48) were assigned to a methylene (HSMQC spectrum). The assignment for the protons was confirmed by the ROESY correlations (see Table 5.1).

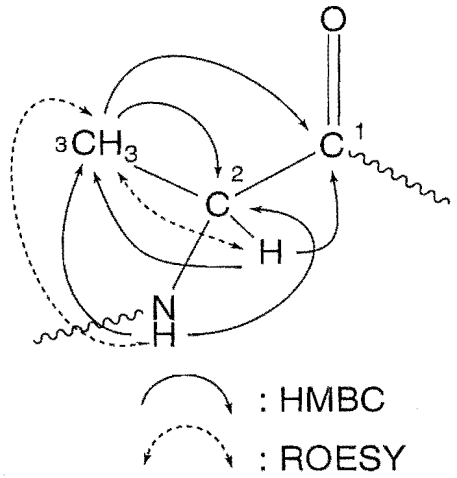


Fig. 5.5.6 Selected HMBC and ROESY correlations of Ala

The spin system of alanine was readily recognised in the TOCSY spectrum because the same AMX system was delineated from both the amide proton at δ 7.32 and the α-proton at δ 4.48. In the HMBC spectrum a two bond correlation to a carbonyl from both the α-proton and an amide proton at δ 8.40 assigned the carbonyl to alanine. The NH of Ala (δ 7.32) was coupled to carbonyl of the adjacent residue (δ 171.3).

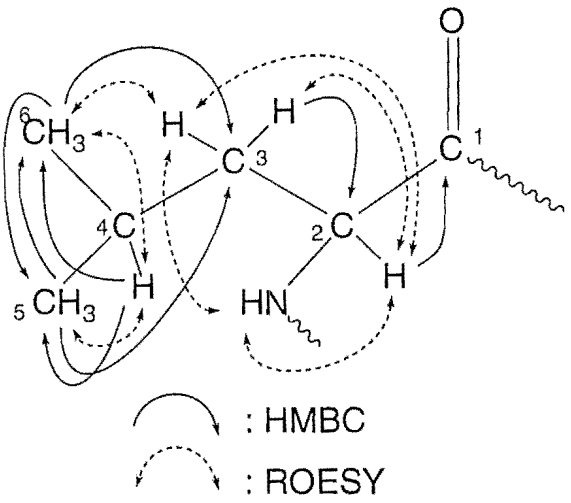


Fig. 5.5.7 Selected HMBC and ROESY correlations of Leu

In the TOCSY spectrum, correlations to δ 1.81, 1.33, 1.0 and 0.96 were observed from the amide proton at δ 8.40. This suggested the existence of a leucine segment. The protons at δ 1.81 and 1.33 were part of a methylene system (both coupled to δ 40.6 in the HSMQC spectrum). However, in the same spectrum, the δ 1.81 proton was also correlated with another carbon at δ 25.5. The DEPT spectrum indicated that the second carbon was a methine, and was assigned to the γ-position of leucine. Despite the overlap of the δ 1.81 proton resonance of the γ-CH and one of the β-methylene protons the correlations extracted from the ROESY spectrum were unambiguous. In the ROESY spectrum, correlations between these two β-methylene protons, between both of them and the α-proton, and between the γ-methine proton and two terminal methyl groups were observed. These results provided supporting evidence for the assignment of this residue as leucine.

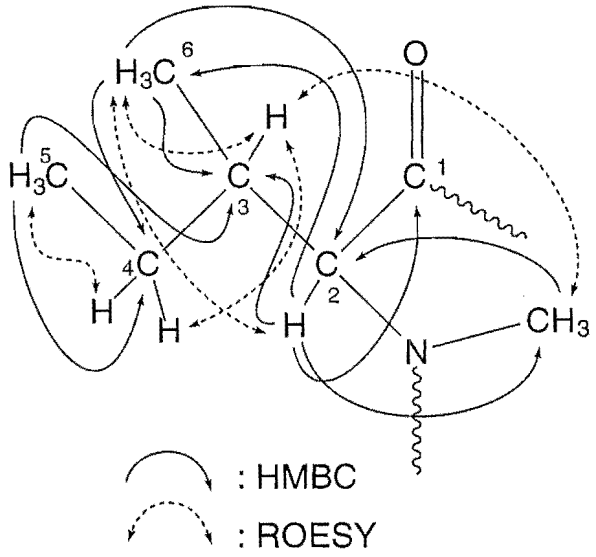


Fig. 5.5.8 Selected HMBC and ROESY correlations of N-MeIle1

The assignment of the two N-methylisoleucine residues are depicted in Figure 5.5.8 and 5.5.9, respectively. Because the chemical shifts were similar and the TOCSY signals heavily overlapped, it was difficult to differentiate between them. Distinction from leucine was based on the observations that the protons of two methyl doublets were correlated to α-carbons at δ 61.1 (αC1) and 55.7 (αC2) in the HMBC spectrum respectively. The corresponding carbon signals for these methyl signals (δ 1.10 and δ 0.80) were located and were assigned to the ε-methyl groups of each N-methylisoleucine. Further examination of HSMQC and HMBC spectra revealed that the αH (δ 5.15) in the first residue was correlated to the carbonyl of valine (2), while the αH in the second residue correlated to the carbonyl of isoleucine and another carbonyl at δ 172.5, and so was assigned to the second N-methylisoleucine. The first residue was designated N-MeIle1, due to its association with Val2.

Extension within the spin system of N-MeIle1 was achieved by further analysis of the HMBC and HSMQC spectra. Both the αH and the ε-methyl group were correlated to a methine carbon at δ 32.5, the β-carbon.

Correlations from the δ- and the ε-methyl protons to a methylene carbon (δ 25.2) located the γ-carbon. The protons attached to these carbons were in turn defined by the HSMQC data. The observation of ROESY correlations between all protons of the N-methylisoleucine confirmed the spin system (see Table 5.1).

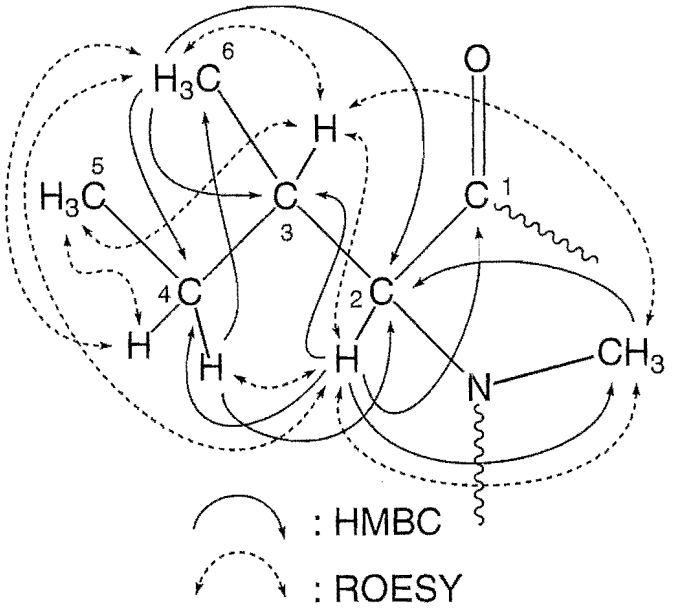


Fig. 5.5.9 Selected HMBC and ROESY correlations of N-MeIle2

In a similar way, the spin system of N-MeIle2 was delineated by HMBC correlations from both the H2 and the ε-methyl protons to the β- and γ-carbons. The δ-carbon was assigned by analogy to IIIe. The following ROESY cross peaks H2/H3, H2/H6, H2/H4, H3/H6, H3/H5, H4/H5 and H6/H4 were recognised, and supported the assignment (see Fig. 5.5.9).

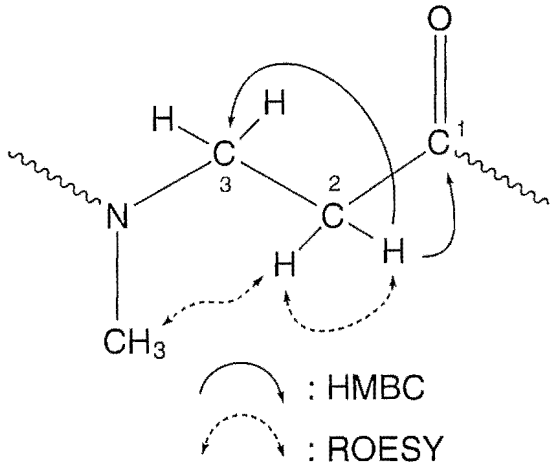


Fig. 5.5.10 Selected HMBC and ROESY correlations of N-Me-β-Ala

The spin system of N-methyl-β-alanine is described next (Fig. 5.5.10). Analysis of the TOCSY spectrum of IIIb failed to identify the residue unambiguously due to the overlapping signals. However, in IIIb one bond H-C couplings from protons at δ 4.63 and 2.73 to a methylene carbon at δ 45.3, and from protons at δ 2.63 and 2.30 to another methylene carbon at δ 34.5 in the HSMQC spectrum unambiguously supported the presence of a N-methyl-β-alanine spin system. The δ 4.63/2.73 protons were designated to β-position due to their appearance at lower field. The large difference between the chemical shift values of these two protons is probably due to the anisotropic effect of the carbonyl group of the adjacent threonine residue. The proton at δ 2.73 must lie above the plan of the carbonyl group. The ROESY correlations between the two α-protons, between the α-proton and the N-methyl protons as well as the amide proton of the adjacent Val2 provided supplementary evidences for this residue. Since a correlation from the β-proton to the carbonyl of threonine was found in the HMBC spectrum, the carbonyl (δ 173.3) that these protons correlated to was attributed to N-methyl-β-alanine.

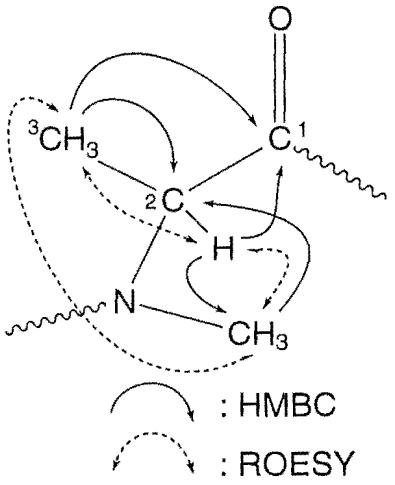


Fig. 5.5.11 Selected HMBC and ROESY correlations of N-MeAla2

Again in contrast to IIIe, the spin system of N-methylalanine (2) was readily defined in the TOCSY spectrum. Confirmation for the partial structure was obtained by ^{1}H-^{1}H decoupling experiment. When the methine signal at δ 5.16 was irradiated, a 3H doublet at δ 1.44 collapsed. The HMBC and ROESY correlations are depicted in figure 5.5.11.

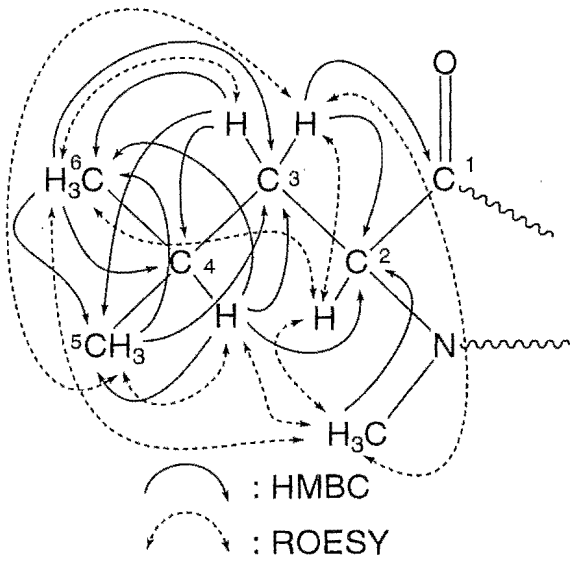


Fig. 5.5.12 Selected HMBC and ROESY correlations of N-MeLeu

The N-methylleucine residue was determined by comparison with the NMR data of theonellapeptolide IIIe. This was followed by a detailed

analysis of the HMBC and ROESY spectra (see Fig. 5.5.12). It is noteworthy that the long-range correlations from both a methyl (N-Me) and from the β-protons to a methine carbon (δ 62.9), defined the α-carbon of N-methylleucine. As the βH was correlated with the carbonyl at δ 172.3 in the HMBC spectrum, this signal was assigned to the N-methylleucine residue.

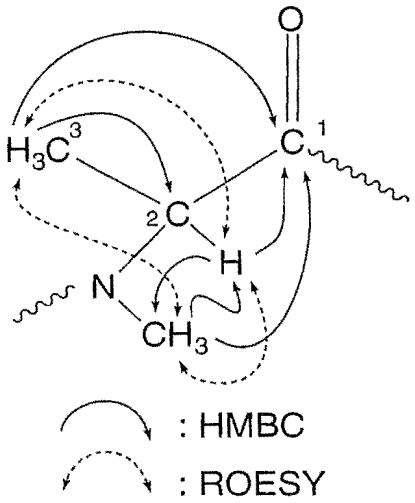


Fig. 5.5.13 Selected HMBC and ROESY correlations of N-MeAla1

Assignments for the last residue, an additional N-methylalanine, are shown in Figure 5.5.13. The TOCSY data failed to provide any clue for this spin system because of overlaps with other signals. In the ^{1}H NMR spectrum, the most likely candidate associated with this residue was an unassigned methyl doublet at δ 1.36. To verify this assumption, ^{1}H-^{1}H decoupling was used. When the δ 1.36 signal was irradiated, a multiplet around δ 5.12 sharpened. When the latter was irradiated, the doublet at δ 1.36 collapsed to a singlet. These two groups of signals were assigned as the α- and β-protons of N-methylalanine. Correlations from the α-proton to the carbonyl of Val1 and a carbon resonance at δ 173.7 in the HMBC spectrum allowed assignment of the δ 173.7 resonance to N-methylalanine (1).

The connectivities between the individual spin systems were based primarily on HMBC correlations from NH, NCH_3 and αH signals to the carbonyl carbon of the adjacent residue. The important correlations include H2 (N-MeAla1)/CO (Val1), H2 (Thr)/CO (N-MeAla1), H3 (N-MeβAla)/CO (Thr), NH (Val2)/CO (N-MeβAla), H2 and N-CH_3 (N-MeIle1)/CO (Val2), NH (βAla)/CO (N-MeIle1), H2 (Ile)/CO (βAla), H2 and N-CH_3 (N-MeIle2)/CO (Ile), H2 and N-CH_3 (N-MeAla2)/CO (N-MeIle2), NH (Ala)/CO (N-MeAla2), NH (Leu)/CO (Ala), and N-CH_3 (N-MeLeu)/CO (Leu). Finally, the H3 of threonine was correlated with the carbonyl of MeLeu, which defined the lactone ring (see Fig. 5.5.14 and Table 5.1). All ROESY correlations, either between the NH and the H2 or between the N-CH_3 and the H2 of the adjacent amino acid moiety were observed (Fig. 5.5.15 and Table 5.1). The only exception was for N-CH_3 of N-MeAla2 to the αH of N-MeIle2. These results confirmed the connectivity established by HMBC analysis.

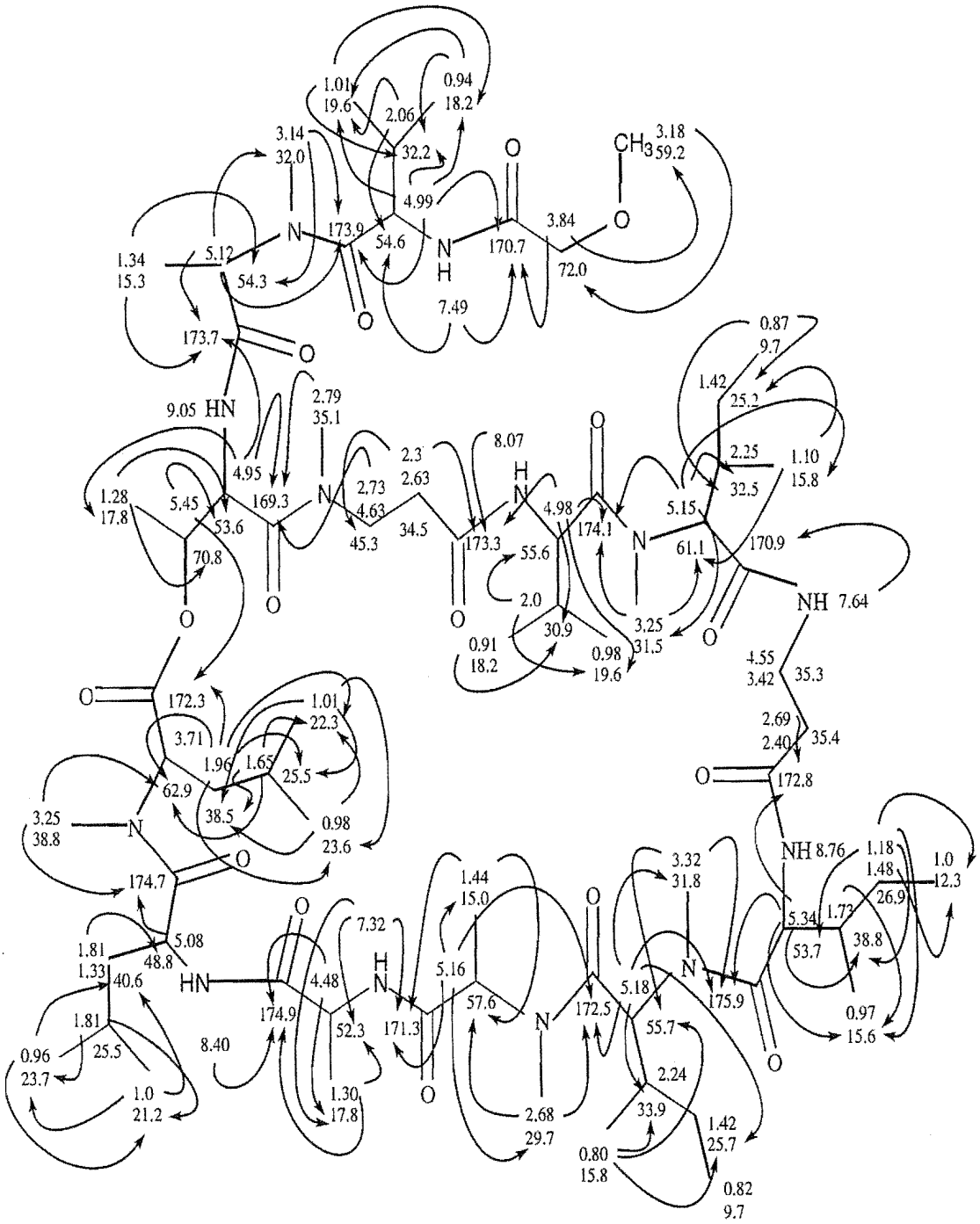


Fig. 5.5.14 HMBC correlations of theonellapeptolide IIIb (**52**)

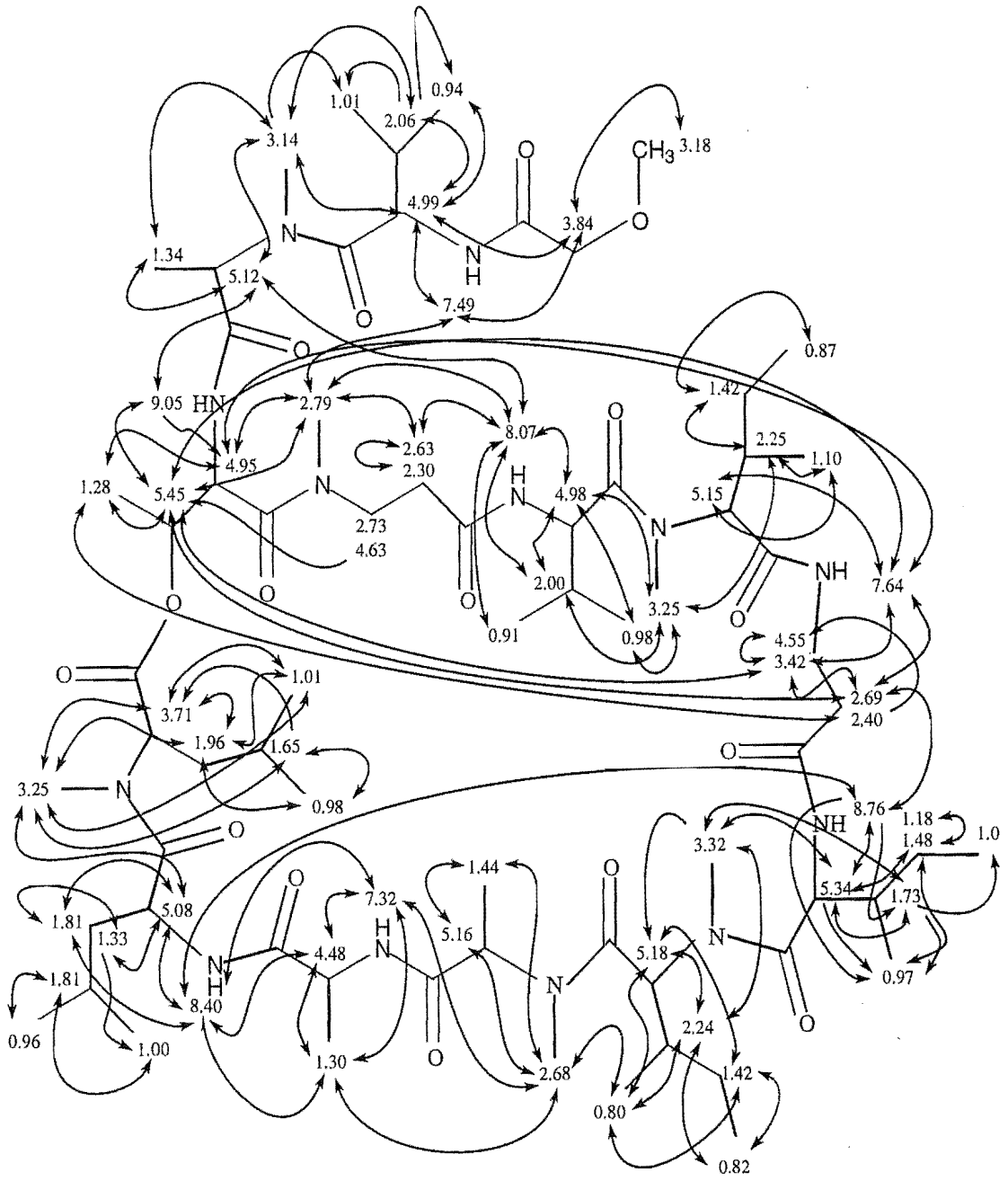


Fig. 5.5.15 ROESY correlations of theonellapeptolide IIIb (**52**)

Table 5.1. Correlated ^{1}H and ^{13}C Spectral Data and ^{1}H-^{1}H NOE Data for Theonellapeptolide IIIb (52)

amino acid		δ[a] ^{13}C (#H)[b]	^{1}H (mult, J(Hz))[c]	HMBC[d]	ROESY[e]
Val (1)	1	173.9 (s)			
	2	54.6 (d)	4.99 (m)	173.9, 170.7, 32.2, 19.6, 17.8	7.49, 3.84, 3.14, 2.06, 0.94
	3	32.2 (d)	2.06 (m)	54.6, 19.6	4.99, 3.14, 1.01, 0.94
	4	17.8 (q)	0.94	32.2, 19.6	4.99
	5	19.6 (q)	1.01	32.2, 17.8	
	NH		7.49 (d, 9.3)	170.7, 54.6	4.99, 3.84, 2.79
MeAla (2)	1	173.7 (s)			
	2	54.3 (d)	5.12 (m)	173.9, 173.7, 32.0	9.05, 8.07, 3.14, 1.34
	3	15.3 (q)	1.34 (d, 7.2)	173.7, 54.3	5.12, 3.14
	NCH_3	32.0 (q)	3.14 (s)	173.9, 54.3	5.12, 4.99, 2.06, 1.34, 1.01
Thr (3)	1	169.3 (s)			
	2	53.6 (d)	4.95 (m)	173.7, 169.3, 17.8	7.64, 2.79, 1.28
	3	70.8 (d)	5.45 (dd, 6.3, 5.7)	172.3, 53.6	9.05, 7.64, 3.42, 2.69, 1.28
	4	17.8 (q)	1.28 (d, 5.7)	70.8, 53.6	5.45, 4.95, 2.40
	NH		9.05 (d, 6.3)		5.45, 5.12, 4.95
MeβAla (4)	1	173.3 (s)			
	2	34.5 (t)	2.63 (m)		8.07, 2.79, 2.30
			2.30 (m)	173.3, 45.3	2.63
	3	45.3 (t)	4.63 (m)		5.45
			2.73 (m)	169.3	
	NCH_3	35.1 (q)	2.79 (s)	169.3	8.07, 7.49, 5.45, 4.95, 2.63
Val (5)	1	174.1 (s)			
	2	55.6 (d)	4.98 (m)	174.1, 173.3, 30.9, 19.6	8.07, 3.25, 2.00, 0.98
	3	30.9 (d)	2.00 (m)	55.6, 19.6	8.07, 4.98, 3.25
	4	18.2 (q)	0.91	30.9	8.07
	5	19.6 (q)	0.98		4.98, 3.25
	NH		8.07 (d, 8.1)	173.3	5.12, 4.98, 2.79, 2.63, 2.00, 0.91
MeIle (6)	1	170.9 (s)			
	2	61.1 (d)	5.15 (m)	174.1, 32.5, 31.5, 15.8	7.654, 1.10
	3	32.5 (d)	2.25 (m)		3.25, 1.42, 1.10
	4	9.7 (q)	0.87 (t, 7.8)	32.5, 25.2	1.42
	5	25.2 (t)	1.42 (m)		2.25, 0.87
	6	15.8 (q)	1.10 (d, 6.6)	61.1, 32.5, 25.2	5.15, 2.25
	NCH_3	31.5 (q)	3.25 (s)	174.1, 61.1	4.98, 2.25, 2.00, 0.98
β-Ala (7)	1	172.8 (s)			
	2	35.4 (t)	2.69 (m)	172.8	8.76, 7.64, 5.45, 3.42
			2.40 (m)		4.55, 1.28
	3	35.3 (t)	4.55 (m)		3.42
			3.42 (m)		7.64, 5.45, 4.55, 2.69
	NH		7.64 (d, 6.6)	170.9	5.45, 5.15, 4.95, 3.42, 2.69
allo-Ile (8)	1	175.9 (s)			
	2	53.7 (d)	5.34 (m)	175.9, 172.8, 38.8, 15.6	8.76, 3.32, 1.73, 1.48, 0.97
	3	38.8 (d)	1.73 (m)	15.6	5.34, 3.32, 1.00, 0.97
	4	12.3 (q)	1.00		
	5	26.9 (t)	1.48 (m)	12.3	5.34, 1.18, 0.97
			1.18 (m)	53.7, 38.8, 15.6, 12.3	1.48
	6	15.6 (q)	0.97		1.48
	NH		8.76 (d, 8.4)		8.40, 5.34, 2.69, 1.73, 0.97
MeIle (9)	1	172.5 (s)			
	2	55.7 (d)	5.18 (m)	175.9, 172.5, 33.9, 31.8, 25.7	2.24, 1.42, 0.80
	3	33.9 (d)	2.24 (m)		5.18, 3.32, 0.82, 0.80
	4	9.7 (q)	0.82 (t, 7.2)		2.24, 1.42
	5	25.7 (t)	1.42 (m)		5.18, 0.82, 0.80
	6	15.8 (q)	0.80 (d, 6.3)	55.7, 33.9, 25.7	5.18, 2.68, 2.24, 1.42
	NCH_3	31.8 (q)	3.32 (s)	175.9, 55.7	5.34, 5.18, 2.24, 1.73
MeAla (10)	1	171.3 (s)			
	2	57.6 (d)	5.16 (m)	172.5, 171.3, 29.7, 15.0	2.68, 1.44
	3	15.0	1.44 (d, 6.3)	171.3, 57.6	5.16, 2.68
	NCH_3	29.7 (q)	2.68 (s)	172.5, 57.6	7.32, 5.16, 1.44, 1.30,. 0.80
Ala (11)	1	174.9 (s)			
	2	52.3 (d)	4.48 (m)	174.9, 17.8	8.40, 7.32, 1.30
	3	17.8 (q)	1.30 (d, 6.0)	174.9, 52.3	8.40, 7.32, 4.48, 2.68
	NH		7.32 (d, 5.7)	171.3, 52.3, 17.8	8.40, 4.48, 2.68, 1.30
Leu (12)	1	174.7 (s)			
	2	48.8 (d)	5.08 (m)	174.7	8.40, 3.25, 1.81, 1.33
	3	40.6 (t)	1.81 (m)	48.8	8.40, 5.08, 1.33
			1.33 (m)		5.08, 1.81, 1.00
	4	25.5 (d)	1.81 (m)	23.7, 21.2	1.00, 0.96
	5	21.2 (q)	1.00	40.6, 23.7	1.81
	6	23.7 (q)	0.96	40.6, 21.2	1.81
	NH		8.40 (d, 8.4)	174.9	8.76, 7.32, 5.08, 4.48, 1.81, 1.30
MeLeu (13)	1	172.3 (s)			
	2	62.9 (d)	3.71 (m)		3.25, 1.96, 1.01
	3	38.5 (t)	1.96 (m)	172.3, 62.9, 25.5, 23.6, 22.3	3.71, 3.25, 1.01, 0.98
	4	25.5 (d)	1.65 (m)	62.9, 38.5, 23.6, 22.3	3.25, 0.98
	5	23.6 (q)	0.98	38.5, 22.3	1.96, 1.65
	6	22.3 (q)	1.01	38.5, 25.5, 23.6	3.71, 3.25, 1.96
	NCH_3	38.8 (q)	3.25 (s)	174.7, 62.9	5.08, 3.71, 1.96, 1.65, 1.01
MeO-CH_2CO	1	170.7 (s)			
	2	72.0 (t)	3.84 (s)	170.7, 59.2	7.49, 4.99, 3.18
	OCH_3	59.2 (q)	3.18 (S)	72.0	3.84

[a] Referenced to residual solvent C_6D_6 δ_H=7.27, δ_C=128.4. [b] ^{13}C spectra recorded on a Varian XL300 at 75 MHz. Number of attached H determined by DEPT. [c] ^{1}H spectra recorded on a Varian Unity 300 at 300 MHz. [d] J_{nxh}=8.3,5.0 arrayed mixing times. [e] A mixing time of 0.15s was used.

5.6 Stereochemistry of Theonellapeptolide IIIb (52)

The solution conformation of theonellapeptolide IIIb (**52**) measured in 30% CD_3OD-C_6D_6 was deduced from a detailed analysis of the ROESY spectra. Also, the conformation was compared with that ascertained for IIIe. The NMR spectra of theonellapeptolide IIIb (**52**) indicated that essentially only a single conformer was present in solution in 30% CD_3OD-C_6D_6 at room temperature. ROESY correlations were observed for H2 (Thr)/NH (βAla), H3 (Thr)/NH (βAla), H3 (Thr)/H2 (βAla), H3 (Thr)/H3 (βAla) and H4 (Thr)/H2 (βAla) as well as NH (Ile)/NH (Leu) (see Fig. 5.5.15), implying that the more hindered side chains all point in the same direction. Since a similar correlation was found in the ROESY spectra of IIIe, it can be imagined that the peptide backbone of IIIb is also folded into the taco-shell shape described earlier in which all the larger residues are aligned along the edge of the taco-shell, while the smaller ones, such as Ala, MeAla and MeβAla are located at the bends.

The ROESY spectrum also showed cross peaks between NH-Leu/NH-Ala and between NH-Ala/NCH_3-MeAla, indicating that the NH of Leu was oriented parallel to the N-H bond of Ala, and similarly for the NH of Ala to the N-CH_3 of MeAla. Conformation of a *cis*-amide bond between residue 9 (N-MeIle2) and 10 (N-MeAla2) was not detected as it was not possible to observe an NOE between the αHs due to their similar chemical shifts (δ 5.18, 5.16).

Correlations between the NH of Val1 and the NCH_3 of MeβAla1, as well as between the αH of MeAla1 and the NH of Val2, are probably a consequence of hydrogen bonding between Val1 and Val2, which would

restrict the motion of the side chain. In summary, these data suggested that IIIb had a similar solution conformation to that of IIIe.

The assignment of the absolute configuration was carried out as follows (Derivatization of the hydrolysate and standard amino acids, and LC/MS work were done by Dr. Shegi and Dr. Pannell at National Institute of Health, U.S.A.). Five pairs of optically pure N-methyl amino acids were prepared from corresponding amino acid enantiomers (see Experimental). Both diastereoisomers of the amino acids were formed by reaction of both the standard amino acids and the acid hydrolysate of the theonellapeptolide IIIb with 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl isothiocyanate (GITC), and Nα-(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA), followed by LC/MS respectively. The derivatives were analysed by LC/MS and the individual amino acid was identified by comparison with both the mass and retention time of the authentic compounds. The component amino acids were determined to be L-Val, D-N-MeAla, L-Thr, D-Val, two residues of L-N-MeIle, D-*allo*-Ile, L-N-MeAla, L-Ala, D-Leu, D-N-MeLeu.

It should be noted that the LC/MS can not distinguish L-N-MeIle from L-*allo*-N-MeIle. The assignment was made as L-N-MeIle by considering that all the known theonellapeptolides contain the L-N-MeIle and that the L-configuration of isoleucine is the predominant form of the enantiomers. L-Val, D-N-MeAla were arbitrarily put on the side chain of the peptolide by analogy to theonellapeptolide IIIe. Incorporation of these stereochemical features into the planar structure of theonellapeptolide IIIb yields its complete structure as Figure 5.5.16.

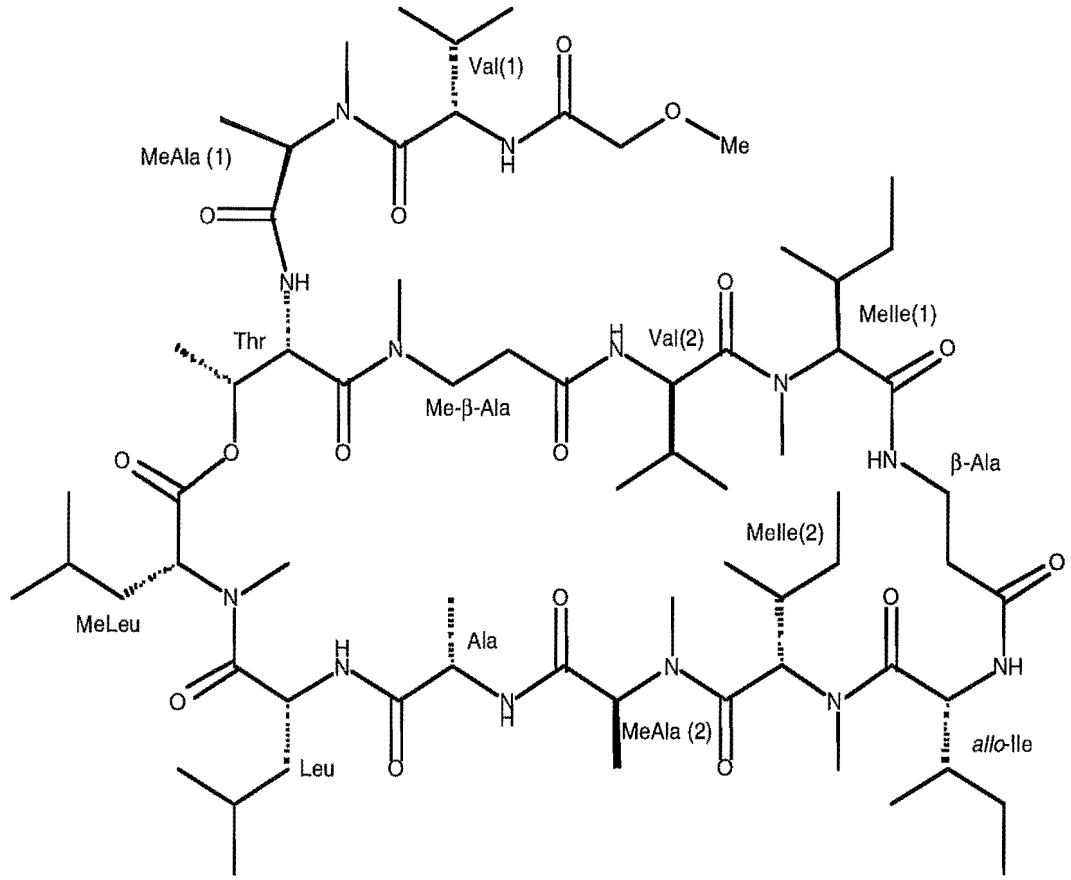


Fig. 5.5.6 Configuration of theonellapeptolide IIIb (**52**)

Chapter Six

Theonellapeptolides IIIa, c, d and Other Peptides

6.1 Introduction

Theonellapeptolides IIIa, c, d and three other peptide compounds were purified by RPHPLC (cf. Section 5.1). The eighth peptide, a very minor component, was only recognised in the purification stage. The Rf value for this peptide was greater than that for IIId. This compound was temporarily designated as theonellapeptolide IIIh. This chapter will discuss the structural elucidation of theonellapeptolides IIIa, c and d using GC/MS, chiral LC/MS, FABMS/MS, and 1D and 2D NMR techniques. Following this, a brief description of the three other peptides (IIIf, g and h) will be given.

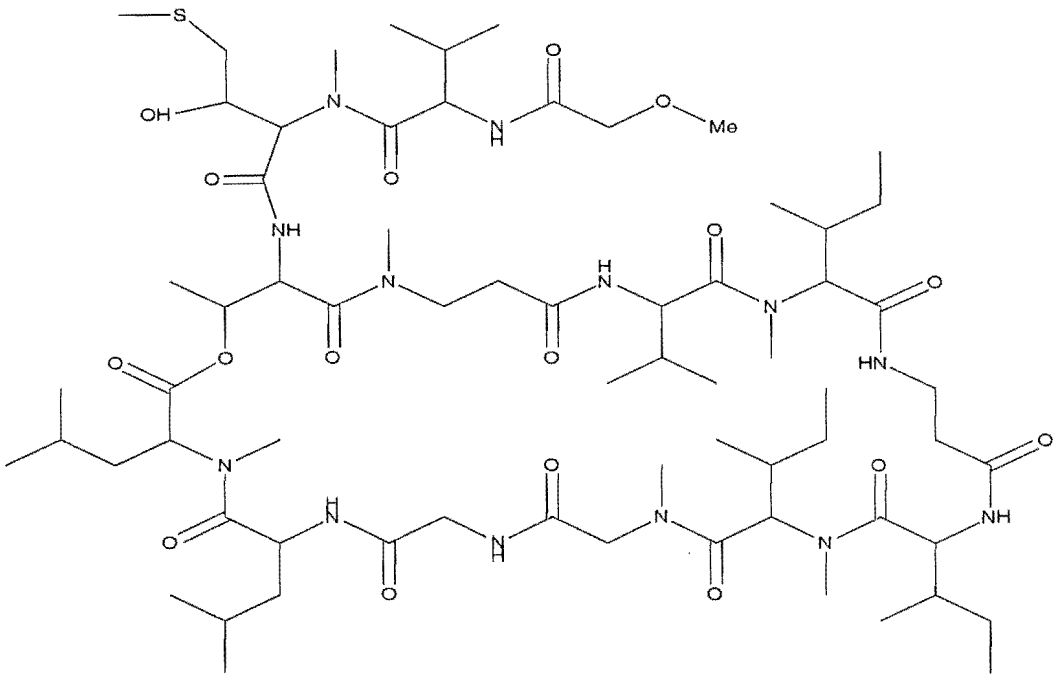
Theonellapeptolides IIIa, c and d have similar ^{1}H NMR spectra to those of IIIb and e. The amide proton resonances appeared between δ 6.6-9.1, N- or O-methyl signals between δ 2.35-3.32, and a sharp methylene peak around δ 3.8. This data implied structural similarities between these compounds. Negative results to ninhydrin were observed for IIIa, c and d, which suggested that the N-terminus of these peptides was also blocked. The IR spectra of them all showed a lactone absorption at 1734 cm^{-1}, indicating peptolide features. Like IIIb and e, the N-terminus in each of these peptolides was probably blocked by a methoxyl acetate moiety. However, the HPLC traces and NMR spectra of these compounds also showed some different characters from those for IIIb and e. For example, after further purification on HPLC, the individual peaks were tested for purity on an analytical HPLC column. The peaks corresponding to IIIa, c and d each showed a shoulder. This behaviour led to confusion on the state of purity of each peptide. When this shoulder was shaved off by repeated chromatography, the ratio had not varied obviously. It was also noticed that the ratio of the shoulder was related to

the length of time standing in the HPLC solvent. Thus, the split peaks were assumed to be caused by different stable conformations of the molecules, rather than impurities. This assumption was confirmed by a series of NMR experiments. Both the ^{1}H and the ^{13}C NMR spectra of these compounds exhibited complex patterns. The ratio of the peaks varied with NMR solvents and concentrations. This feature caused problems in the assignment of the spectra. The NMR spectra were analysed using comparisons of the spectra obtained in different solvent with those of IIIb and e.

6.2 Theonellapeptolide IIIa

Theonellapeptolide IIIa (**54**), a colourless material drying to a white non-crystalline powder, is optically active. The molecular weight of 1584.8100 $(M+Cs)^+$, obtained from HRFABMS, is 34 amu higher than that of theonellapeptolide IIIe. This mass difference suggested that either a chlorine atom had replaced one of the hydrogen atoms in IIIe or possibly an oxygen plus a sulphur had substituted for a CH_2 to give IIIa.

The ratio of the M^+/M^++2 peaks was difficult to interpret due to the inherent complexity in the vicinity of the molecular ion in a compound of this mass. The analysis of the isotope pattern in the vicinity of the molecular ion did not strongly support the presence of chlorine in the molecule. Based on ^1H and ^{13}C NMR data (no double bond was observed in spectra of IIIa), results of the amino acid analysis (13 amino acids were determined), as well as the number of double bond equivalents, a molecular formula $C_{70}H_{125}O_{17}N_{13}S$ is that most approximate for IIIa. The ^1H NMR spectrum (measured in 35% CD_3OD/C_6D_6, Fig. 6.2.1) showed six amide signals (δ 6.78-8.11). However, seven resonances in this area were recognised when the spectrum was obtained in 10% H_2O/CD_3OD. This implied that there were two signals overlapped in the former spectrum. A sharp singlet at δ 3.71 is characteristic of the methylene protons of methoxy-acetate, as contained in theonellapeptolides IIIb and e. Seven three-proton singlets between δ 2.38 and 3.20 can be attributed to six N-methyl groups and a methoxy group. Peak splitting was a common phenomenon in the ^{13}C spectrum due to existence of several conformations in the solvent. Careful examination of the ^{13}C NMR spectrum resulted in the recognition of 12 carbonyl signals recognised between δ 168.2 and 174.9. Obviously some overlaps existed.



Theonellapeptolide IIIa (**54**)

Amino acid analysis by GC/MS revealed the presence of Val (2), Thr (1), N-methyl-leucine (N-MeLeu) (1), N-methyl-isoleucine (N-MeIle) (2), Leu (1), alle (1), Ala (1), β-Ala (1), N-methyl-alanine (N-MeAla) (1) and N-methyl-β-alanine (N-MeβAla) (1) residues. A comparison with the GC/MS of the derivatives of the hydrolysate from IIIe indicated that all but one of the component amino acids in these two peptolides were the same. A new peak eluted just before the threonine peak in the GC trace. This chromatographic behaviour implied that the new amino acid was either of similar polarity or similar molecular weight with threonine. The corresponding MS spectrum had a *m/z* 110 ion, which is characteristic of N-methyl amino acids. Another characteristic signal appeared at *m/z* 61 as a base peak in this spectrum. This signal is unique for methionine. It suggested that the residue has a similar partial structure, $-CH_2-S-CH_3$, as found in methionine, but with a branch

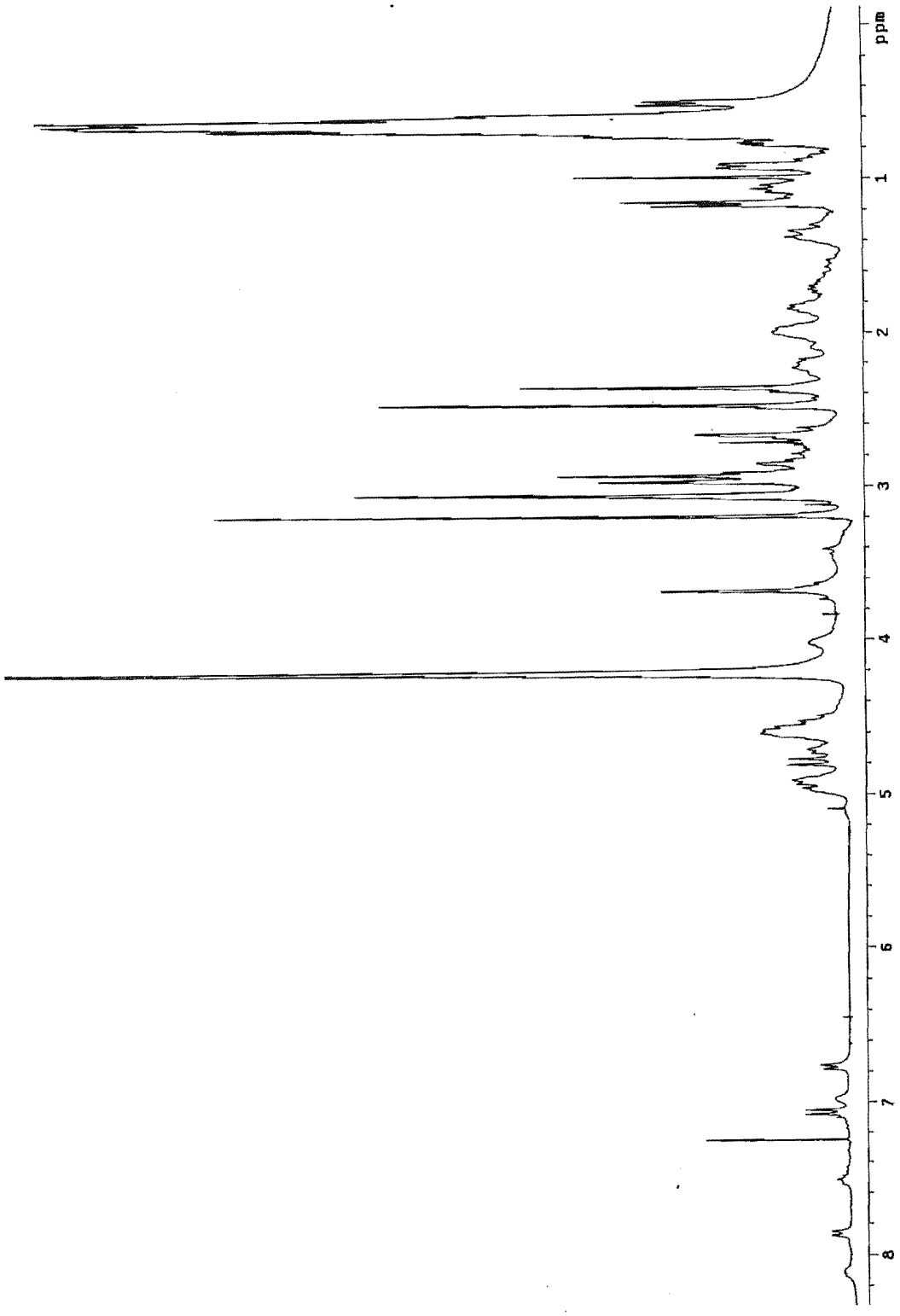
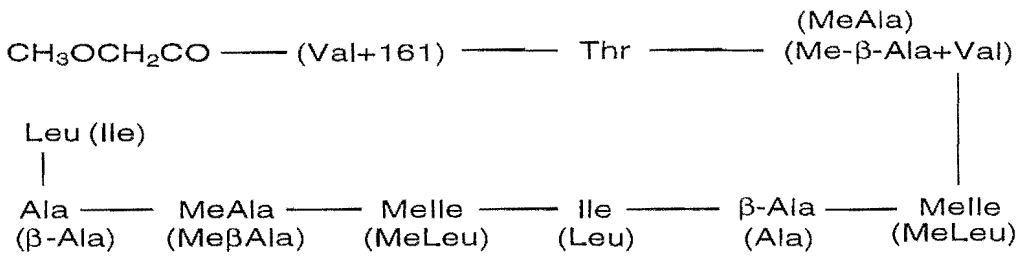


Fig. 6.2.1 ^{1}H NMR spectrum of theonellapeptolide IIIa (**54**)

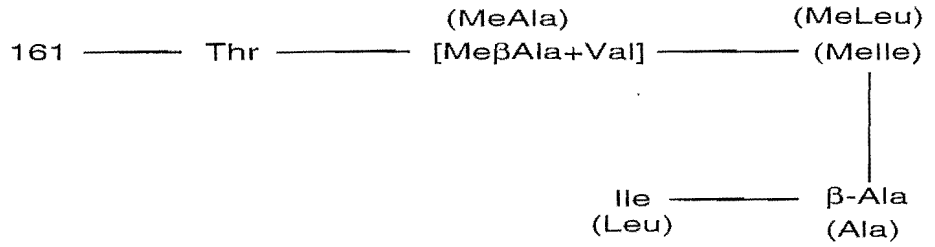
at the β-carbon to make the ion more stable after cleavage of the bond between the β- and γ-carbons. Taking account of the molecular formula of the component, it suggested that it was an hydroxyl group attached to the β-carbon.

This depsipeptide was sequenced by FAB MS/MS analysis after the linear peptide (**55**) was prepared by methanolysis with sodium methoxide. In the CAD spectrum (Fig. 6.2.2) of this ring-opened peptide, two series of ions were recognised. One of these series exhibited a successive loss of amino acid residues from an ion at *m/z* 1051, which arose by elimination of 433 amu from the protonated parent ion. A sequence of 1051-MeβAla-Val-MeIle-(βAla+Ile+MeIle)-MeAla-(Ala+Leu)-MeLeu-OCH_3 can be proposed. The other series can be described as $(M+H)^+$-343-MeAla-MeIle-(Ile+βAla)-(MeIle+Val+MeβAla+Thr)-161-Val-$COCH_2OCH_3$. These two series of ions can be explained by Y_i and B_i types of cleavage, respectively (cf. Fig 4.4.1 in Chapter 4). However, clarification of the ambiguities, such as the composition of the fragments of the *m/z* 434, 344, 160 ions, and the sequences Ile+βAla and MeIle+Val+MeβAla+Thr, required more information. This was achieved by examination of the daughter ion spectrum of $(M+H)^+$ (Fig. 6.2.3), in which the ions were abundant in the higher mass area. Several peaks which had not been recognised in the original spectrum were observed in this link scan. The ions at *m/z* 1325, 1212, 1141, 929, 816 (weak), 745, 618, 434 (weak) and 333 were assignable to B_{12}, B_{11}, B_{10}, B_9, B_8, B_7, B_6, B_5, B_3 and B_2. The peptide fragment **55a** may represent the corresponding sequence.



Peptide fragment **55a** (contents in () imply either isobaric amino acid residue or loss of dipeptide)

In the same spectrum, another series of ions at *m/z* 1313, 1152, 1051, 867, 740 and 556, corresponding to Y_{12}, Y_{11}, Y_{10}, Y_8, Y_7 and Y_5, may be attributed to peptide fragment **55b**.



Peptide fragment **55b** (contents in () imply either isobaric amino acid residue or loss of dipeptide)

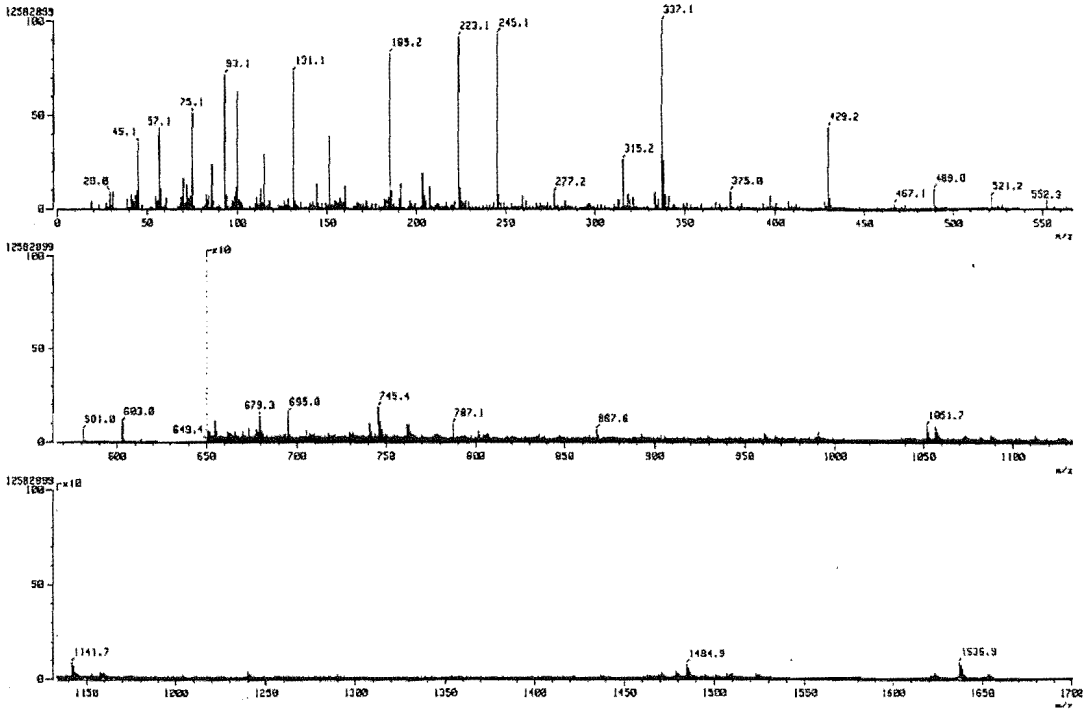


Fig. 6.2.2 FABMS spectrum of compound **55**

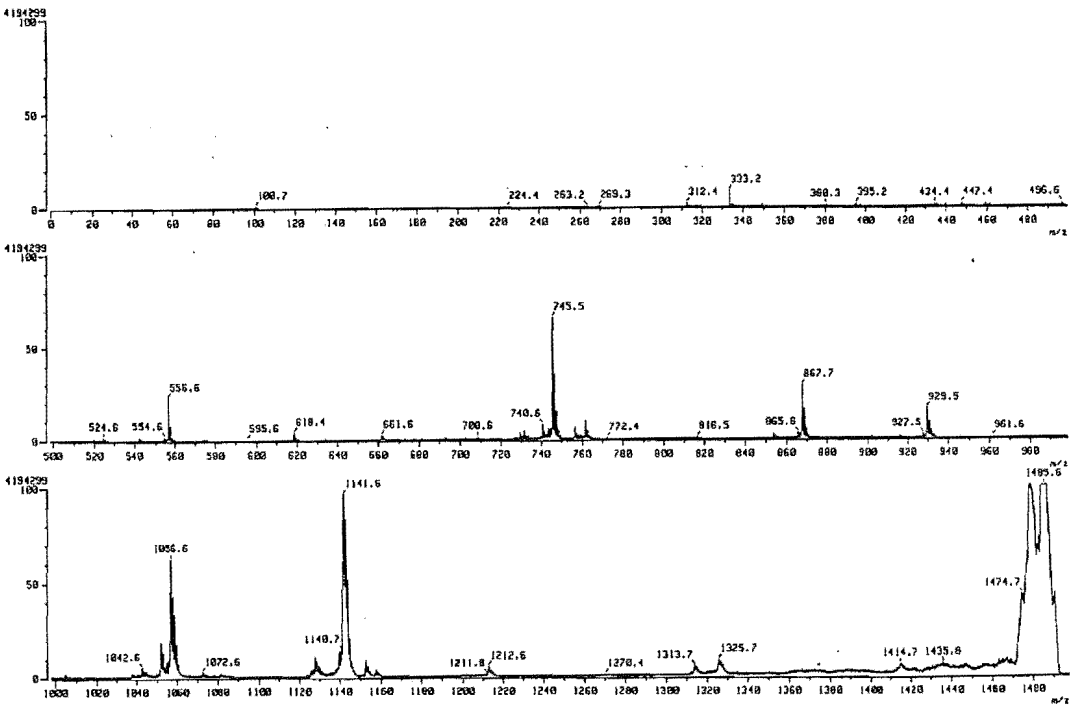


Fig. 6.2.3 Link scan at *m/z* 1485

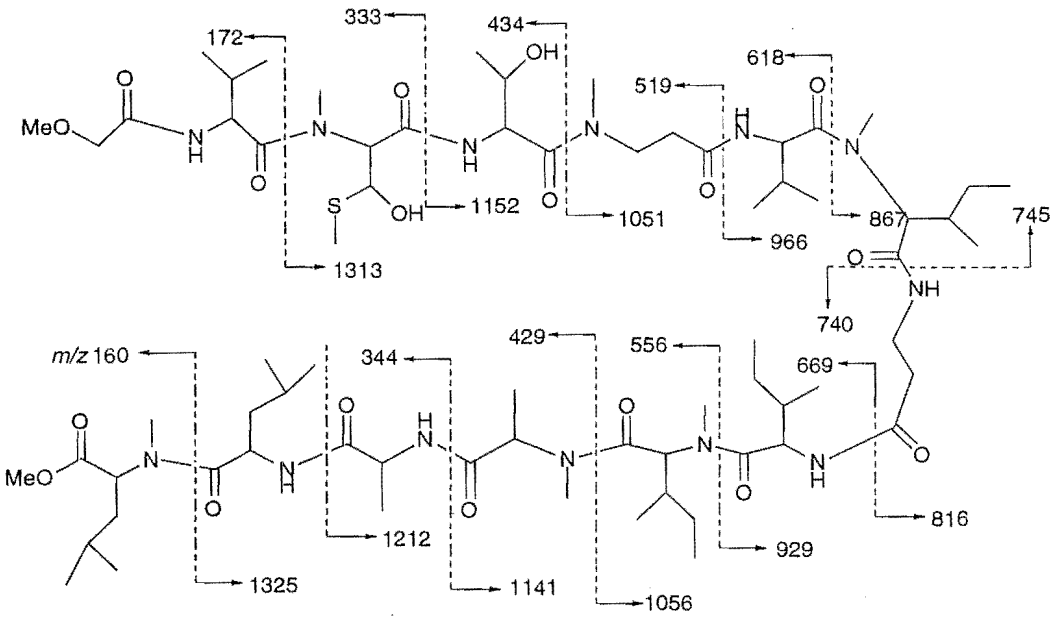


Fig. 6.2.4 FABMS/MS fragmentation ions from both N- and C-termini of the ring-opened peptide (**55**)

A comparison of the fragment **55a** with the pseudo-molecular ion of peptide **55** revealed that there was a 159 amu difference between them. A methyl ester of N-methyl leucine or N-methyl isoleucine at C-terminus accounts for this fragment. So, the tripeptide at the C-terminus can be mapped as MeLeu-Leu-Ala or comprised of their isobaric amino acids. This tripeptide was confirmed by comparing the daughter ion spectrum of *m/z* 343 of **55** with that of ring-opened IIIe (**51**).

A link scan at *m/z* 1051 demonstrated that the ions at *m/z* 966, 867 and 740 are daughter ions, which confirmed that the dipeptide comprising the fourth and fifth residues from the N-terminus was MeβAla-Val. It is noteworthy that a much clearer daughter ion spectrum was obtained when the link scan was performed at three amu off *m/z* 1051 (Fig. 6.2.5). This procedure efficiently avoided overlaps of the daughter ion spectra from the *m/z* 1051 and 1056 ions. When the daughter ion spectrum of *m/z* 1056 (Fig. 6.2.6) was subtracted from that of *m/z* 1051, it showed a series

of product ions formed from B-type cleavage such as B_8, B_7, B_6, B_5, B_4, B_3 and B_2. Although the ion B_7 at *m/z* 816 was relatively weak, it could clearly be identified as a daughter ion from *m/z* 1056 as the corresponding peak was not apparent in the link spectrum from *m/z* 1051 (actually scanned at *m/z* 1048). This offered further evidence for the sequence β-Ala-Ile at the seventh and eighth residues from the N-terminus.

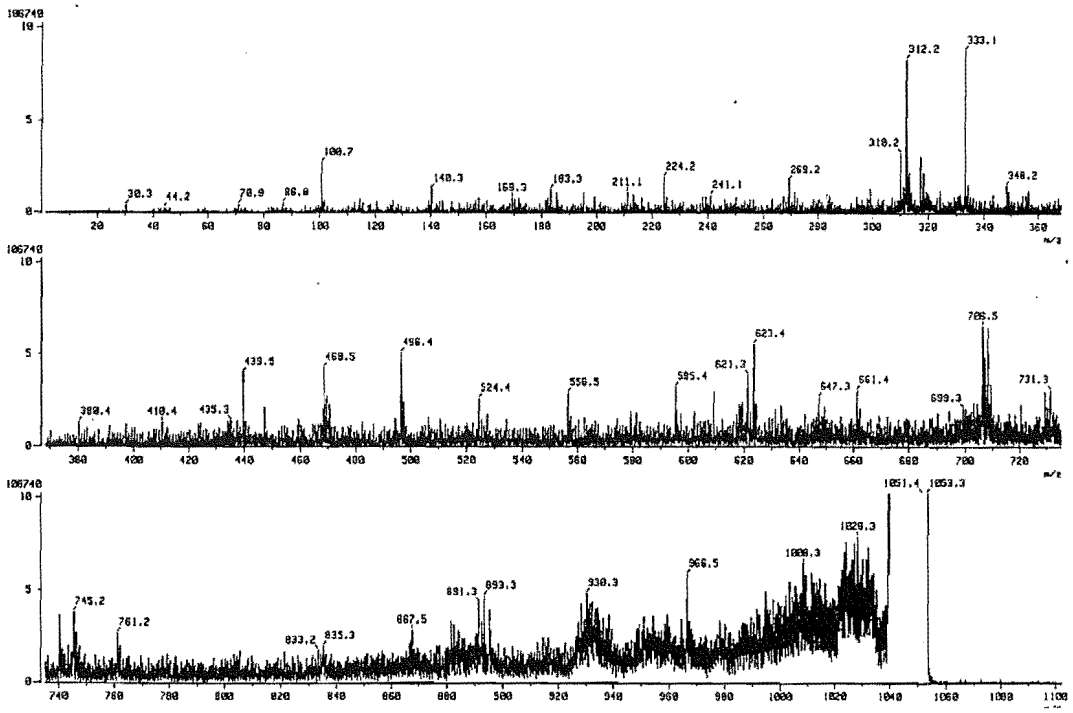


Fig. 6.2.5 Link scan at *m/z* 1048

A mass gap between the N-terminal residue, valine and the third residue, threonine, was recognised in several link-scan spectra. For example, such as scanning at *m/z* 1484 (Fig. 6.2.3), 1313 and 333. The daughter ion spectrum for the ion *m/z* 333 was shown in Fig. 6.2.7. This demonstrated that the daughter ion *m/z* 172 was formed by a loss of 161 amu from the parent ion, while the presence of a *m/z* 162 ion, which may be generated from the same cleavage accompanying by a hydrogen transfer, gave

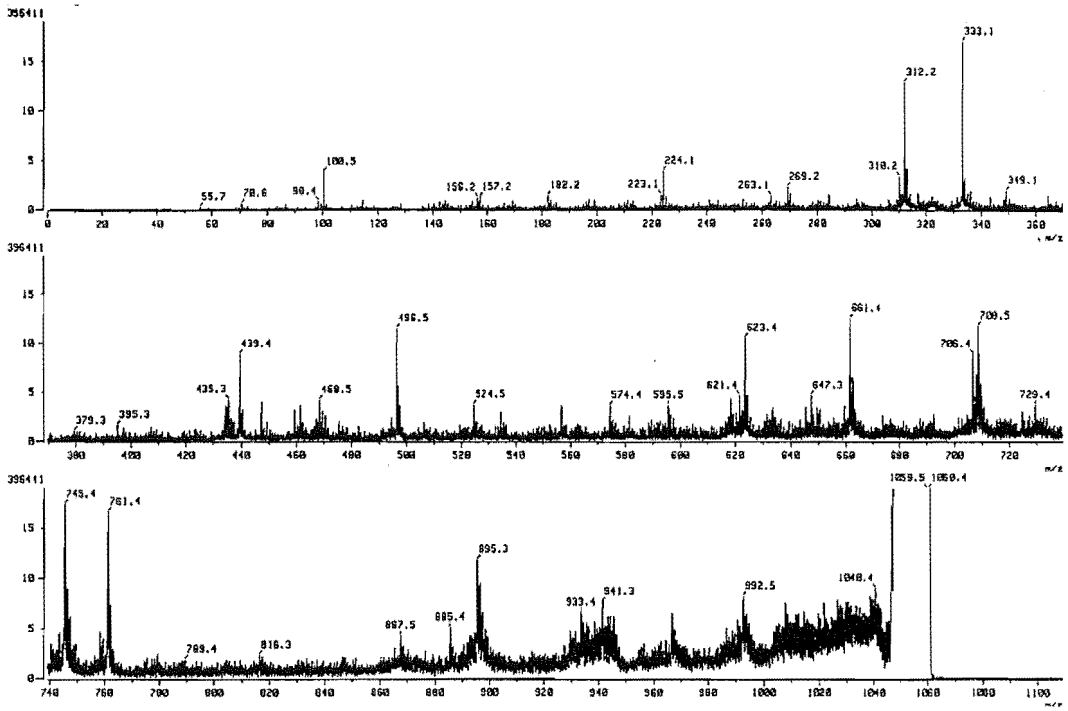


Fig. 6.2.6 Link scan at *m/z* 1056

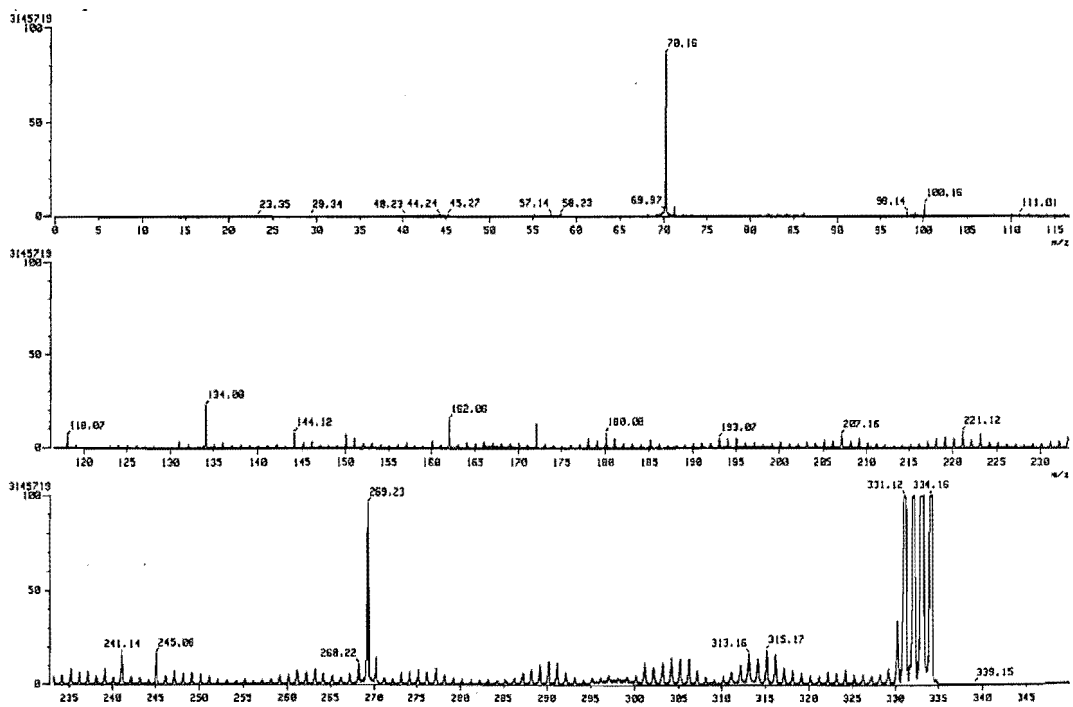
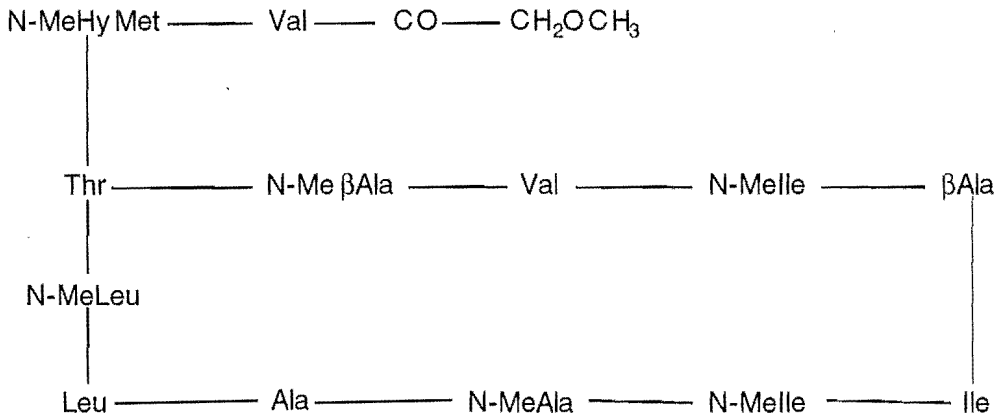


Fig. 6.2.7 Link scan at *m/z* 333

further support to the above assumption. This assignment was also consistent with the mass difference between the intact molecules of peptolides IIIa and IIIe. Hence, the new amino acid was located at the second residue. A strong signal at *m/z* 269 (Fig. 6.2.7) may be explained by successive loss of H_2O (ion at *m/z* 315) and SCH_3 from a hydroxymethionine residue. This provided more evidence for the assignment. Based on the above analysis, the intact sequence of theonellapeptolide IIIa can be described as follows.



Intact sequence of theonellapeptolide IIIa (**54**)

N.B. Ala and βAla, MeAla and MeβAla, Leu, Ile and *allo*-Ile, and MeLeu, MeIle and MeaIle are isobaric and their positions are exchangeable.

Spin systems corresponding to each amino acid residue were mainly recognised by interpretation of the COSY and TOCSY spectra of IIIa. Integration of these data with analysis of the HSMQC and HMBC spectra helped to distinguish between the isobaric amino acids. Particularly important were the HMBC correlations from H6 to C2 of the isoleucine (or N-methyl isoleucine), which differentiated the Ile (N-methyl isoleucine) from leucine (N-methyl leucine). It was not possible to define the spin system of the hydroxy methionine residue as the spectrum was

heavily overlapped in critical regions. The only supporting evidence was a correlation between a methyl signal at δ 2.38 and a methylene carbon at δ 50.6 in the HMBC spectrum.

The intact sequence of the peptolide was consolidated by observation of the following HMBC correlations: H2 (methoxyl acetate) and NH (Val 1)/CO (methoxyl acetate), H2 (Val 1) and N-CH_3 (MeHyMet)/CO (Val 1), H2 (Thr)/CO (MeHyMet), H2 (Thr) and N-CH_3 (MeβAla)/CO (Thr), N-CH_3 (MeβAla)/C3 (MeβAla), H2 (Val 2)/CO (MeβAla), H2 and N-CH_3 (MeIle 1)/CO (Val 2), NH and H2 (Ile)/CO (βAla), H2 (Ile) and N-CH_3 (MeIle 2)/CO (MeIle 2), H2 (MeIle 2) and N-CH_3 (MeAla)/CO (MeIle 2), H2 (MeAla) and NH (Ala)/CO (MeAla), H3 (Ala)/CO (Ala), H2 (Leu) and N-CH_3 (MeLeu)/CO (Leu), N-CH_3 (MeLeu)/C2 (MeLeu), and H3 (Thr)/CO (MeLeu). Failure to observe any correlation between the residues 6 (MeIle 1) and 7 (βAla), and 11 (Ala) and 12 (Leu) resulted in the unsuccessive peptide sequence. However, the intact sequence depicted in Fig. 6.2.4 was the only possible combination, when account was also taken of the MS/MS analysis of the sample.

The absolute configuration of theonellapeptolide IIIa was deduced by analysis of the acid hydrolysate converted to Nα(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA) and 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl isothiocyanate (GITC) derivatives using non-chiral LC/MS. By comparing both the mass and the retention time of each amino acid derivative with those of the authentic (S)- and (R)- amino acids, it was ascertained that L-Val, L-Thr, D-Val, L-N-MeIle, D-*allo*-Ile, L-N-MeAla, L-Ala, D-Leu and D-N-MeLeu, as well as βAla, MeβAla and a new residue, were present in IIIa. The optical properties of the last residue have not been determined. Methyl isoleucine, rather than methyl

allo-isoleucine, was assigned based on an analogy to theonellapeptolide IIIe. On the assumption that the location of L-valine and D-valine residues are the same as in IIIe, the configuration of the theonellapeptolide IIIa can be depicted as in Fig. 6.2.8.

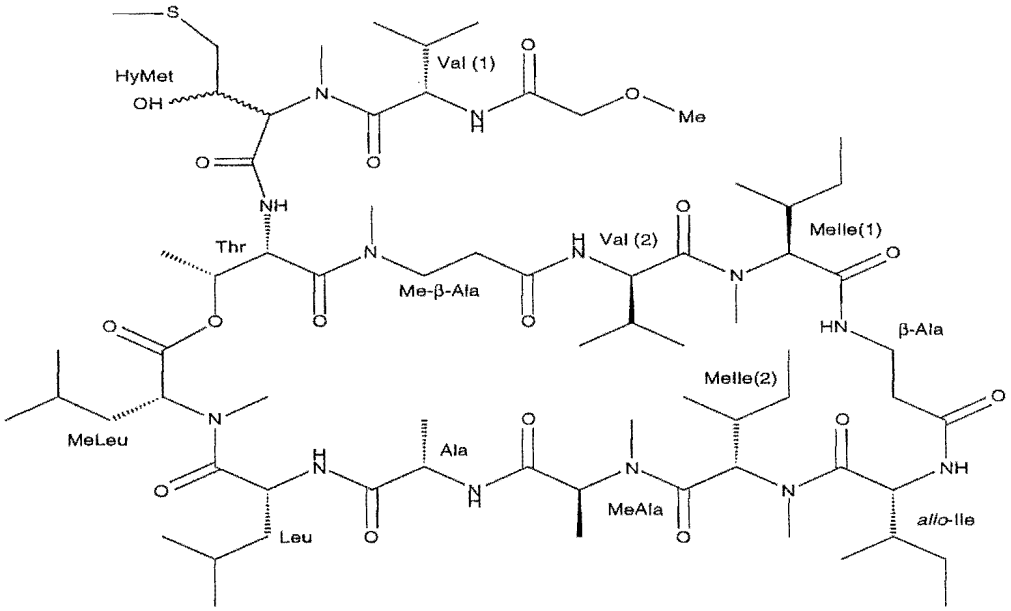
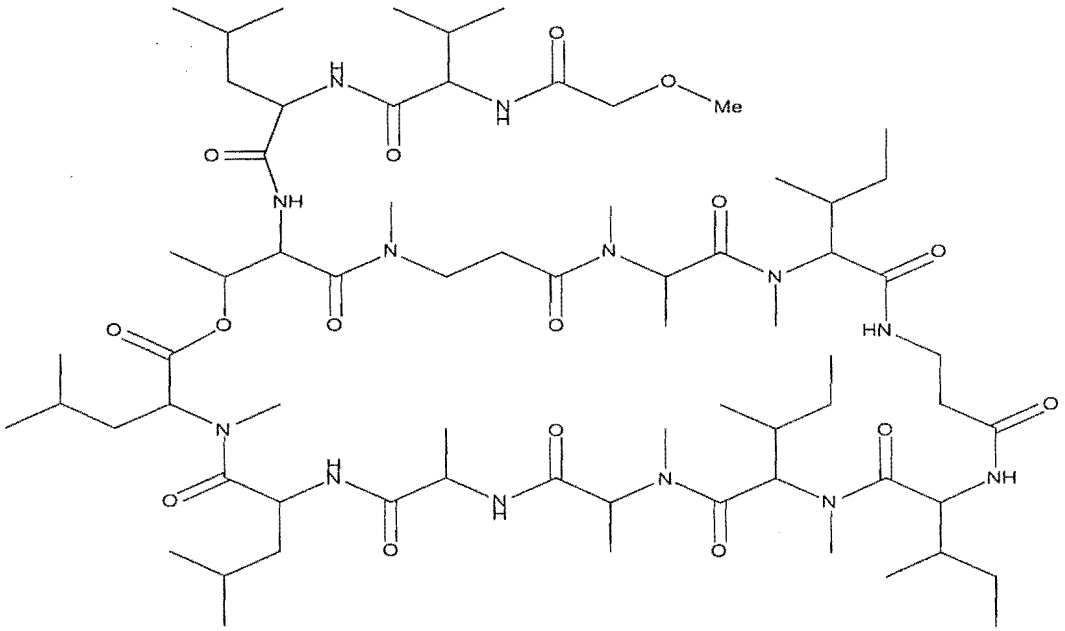


Fig. 6.2.8 Stereochemistry of theonellapeptolide IIIa (**54**)

6.3 Theonellapeptolide IIIc

Theonellapeptolide IIIc (**56**) exhibited cytotoxicity against P338 with an IC_{50} of 11.5 μg/mL. In agreement with HRFABMS data on the $(M+Cs)^+$ ion (1522.8339, Δ+4.8), the molecular formula of IIIc was assigned as $C_{69}H_{123}O_{16}N_{13}Cs$. IIIc appeared as a mixture of conformers in all the NMR solvents tested ($CDCl_3$, CD_3OD, $CD_3OD/CDCl_3$, H_2O/CD_3OD and pyridine-d_6). The NMR spectra were not well resolved. However, the spectral pattern observed was similar to those observed for IIIb and IIIe, suggesting that IIIc shared the same basic skeleton as the other theonellapeptolides. The main conformer in CD_3OD showed seven amide protons (between δ 7.38 and 9.10), six N-methyl, plus one methoxy group (δ 2.58, 2.63, 3.08, 3.14, 3.16 and 3.28) resonances (Fig. 6.3.1). As expected, split peaks caused by different conformers predominated in the ^{13}C NMR spectrum, especially in the carbonyl signal area. IIIc was found to contain the same amino acids as IIIe, but varying in ratios. For instance, two leucines and two N-methyl alanines *vs* one N-methyl leucine



Structure of theonellapeptolide IIIc (**56**)

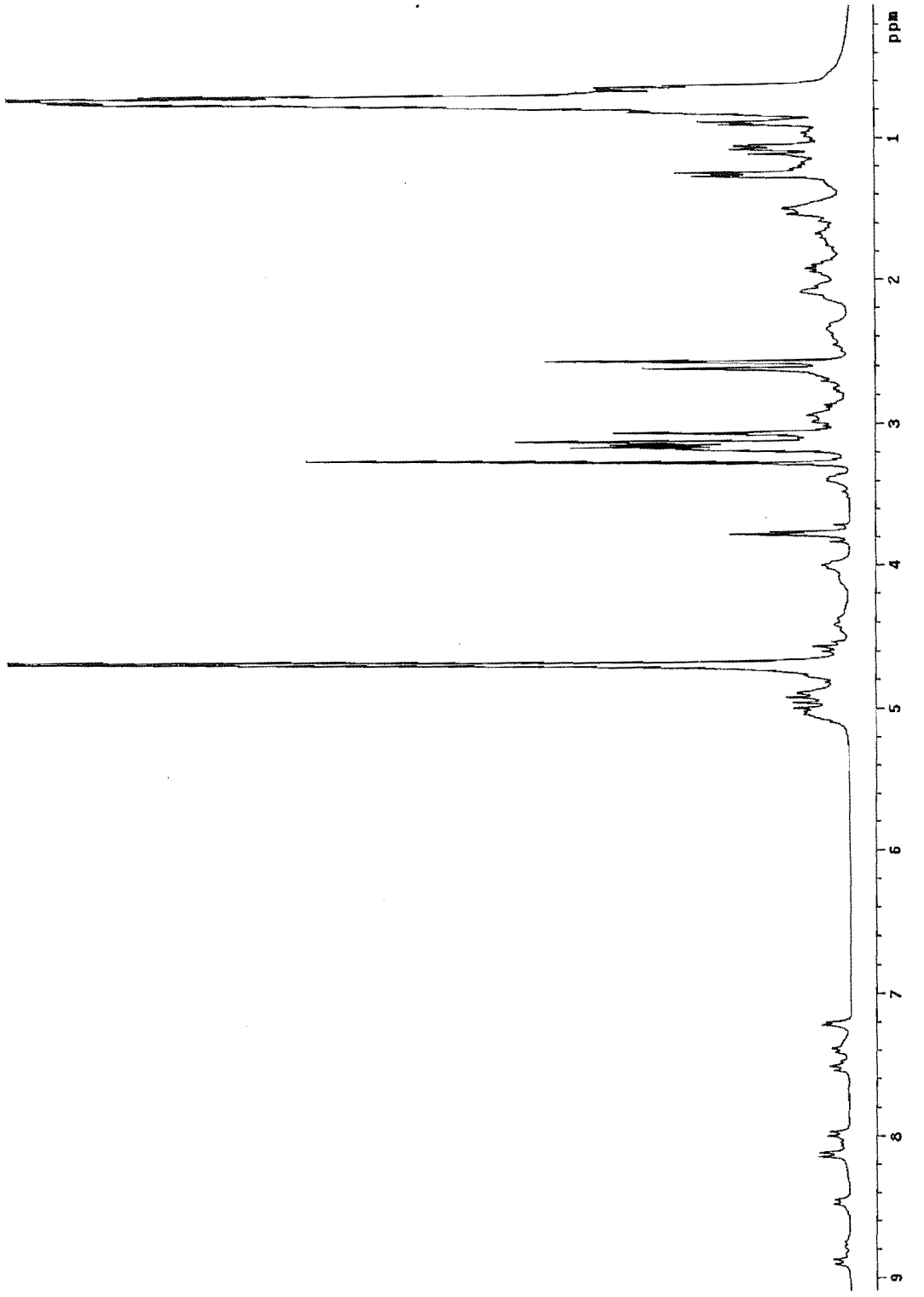
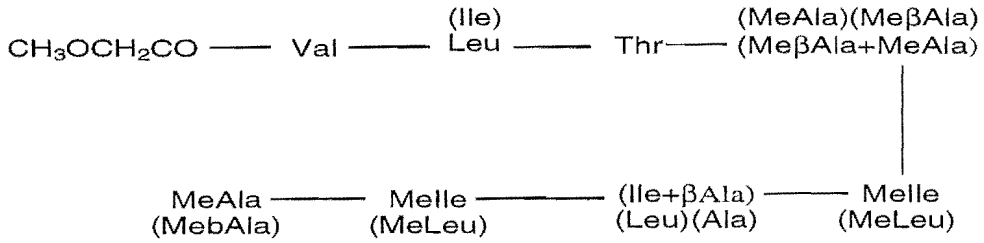


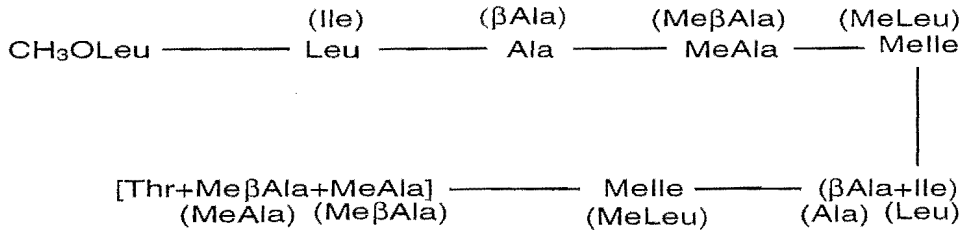
Fig. 6.3.1 ^{1}H NMR spectrum of theonellapeptolide IIIc (**56**)

and one valine.

Figure 6.3.2 shows the FAB/CAD mass spectrum of ring-opened peptide **57**, in which the $(M+H)^+$ ion was observed at *m/z* 1422.9. Two sequences of peptide fragments (**57a** and **57b**), corresponding to successive loss of amino acid residues from C- and N-termini respectively, were proposed by incorporation of the fragmentation information from Fig. 6.3.2 and the daughter ion spectrum of $(M+H)^+$ ion (Fig. 6.3.3).



Peptide fragment **57a**



Peptide fragment **57b**

These sequences were confirmed by a series of link scan experiments. The presence of the leucine unit in the second position from N-terminus was demonstrated by a daughter ion *m/z* 172 from the *m/z* 285 ion. These two ions can be claimed to be the daughter ions of *m/z* 386 since they were both found in the daughter ion spectrum of *m/z* 386 (Fig. 6.3.4). In most of the mass spectra, there was a gap of 170 amu between

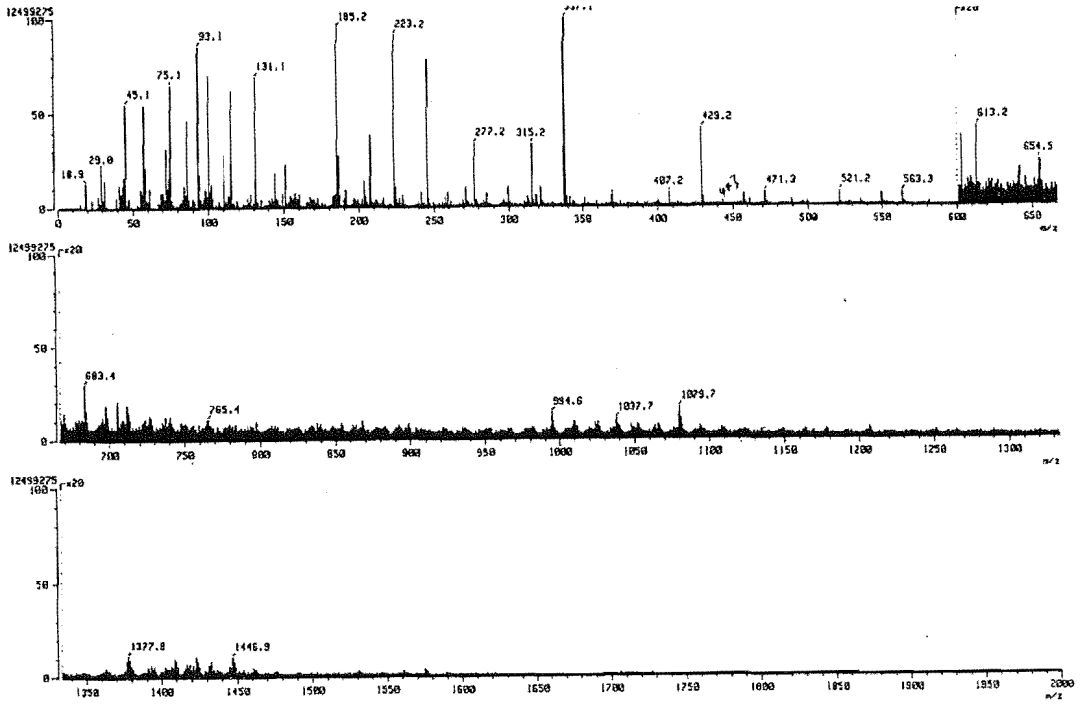


Fig. 6.3.2 FABMS of Compound **57**

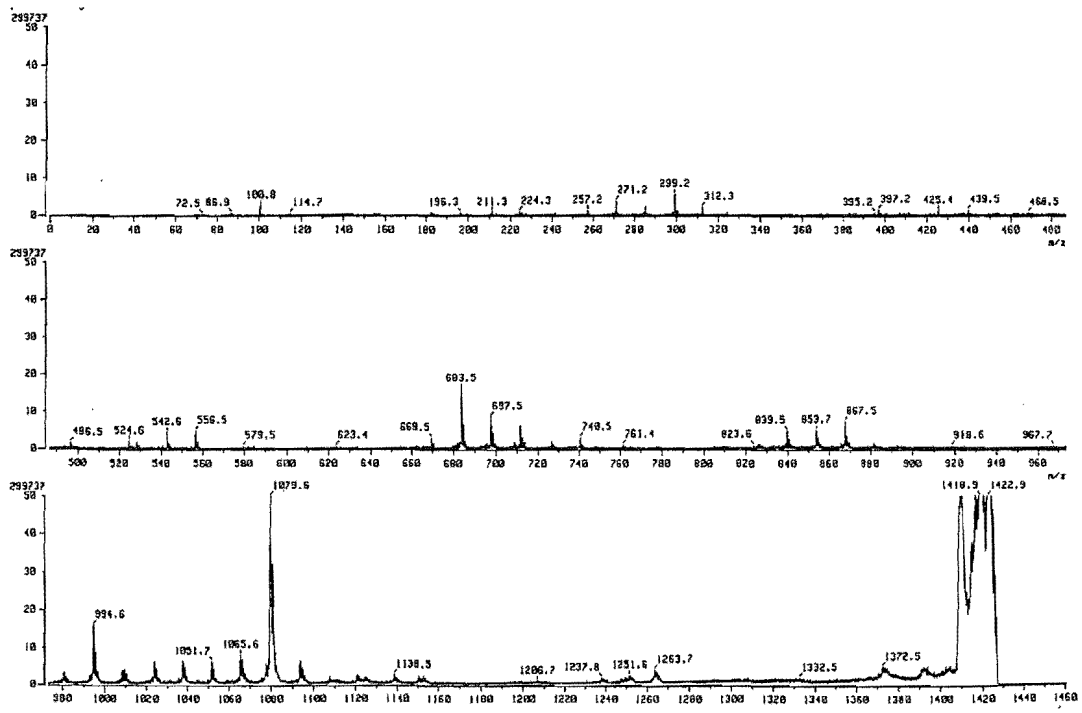


Fig. 6.3.3 Link scan at *m/z* 1423

threonine and methyl isoleucine. This mass difference can be attributed to a dipeptide fragment, which may contain either two methyl alanine (methyl β-alanine) or one valine and one alanine (β-alanine). Evidence supporting the former sequence is that a fragment at *m/z* 471 was observed in the original CAD spectrum. A fragment CH_3OCH_2CO-Val-Leu-Thr-MeβAla (MeAla) can account for the composition of this ion. A link scan on *m/z* 471 yielded a relatively abundant 284 ion, which added greater certainty to the assignment. However, it was not confirmatory because no direct evidence for the origin of the ion *m/z* 471 was found. A link scan from *m/z* 994 (B_9) gave a fragment at *m/z* 271 (Fig. 6.3.5), which was formed from the internal cleavage and can reasonably be explained by a tripeptide fragment comprising Thr-MeβAla-MeAla or Thr-Ala-Val, while a coexisting internal ion at *m/z* 299 has only one possible composition, Leu-Thr-MeβAla. This provided indirect supporting evidence for the MeβAla-MeAla sequence. The NMR data also offered some clues for this arrangement. The ^1H NMR spectrum showed that IIIc had seven amide protons. (If the peptide has the sequence, Ala-Val, the number of the amide protons will be more than seven.) In the linked spectrum from *m/z* 1079, the presence of ions at *m/z* 867 and 796 (Fig. 6.3.6) established the order of the dipeptide, β-Ala-Ile. This sequence can also be traced through the *m/z* 312, 241 and 128 ions when the *m/z* 867 ion was scanned. Thus, a whole sequence for IIIc, based on Fragments **57a** and **57b**, can be constructed (Fig. 6.3.7).

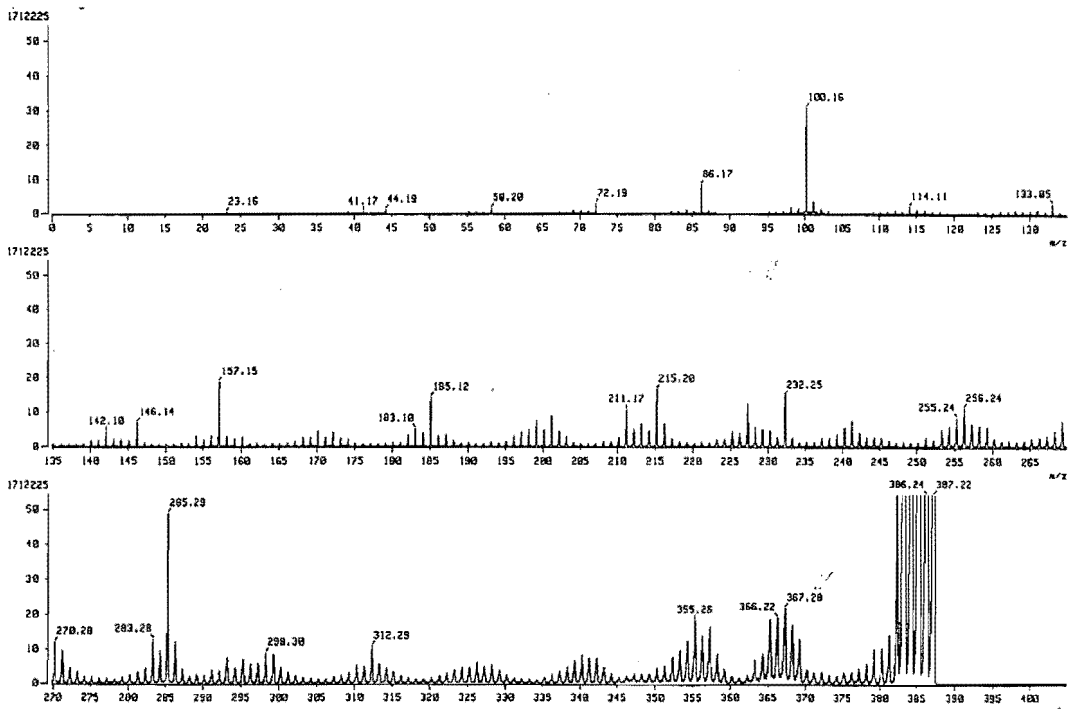


Fig. 6.3.4 Link scan at *m/z* 386

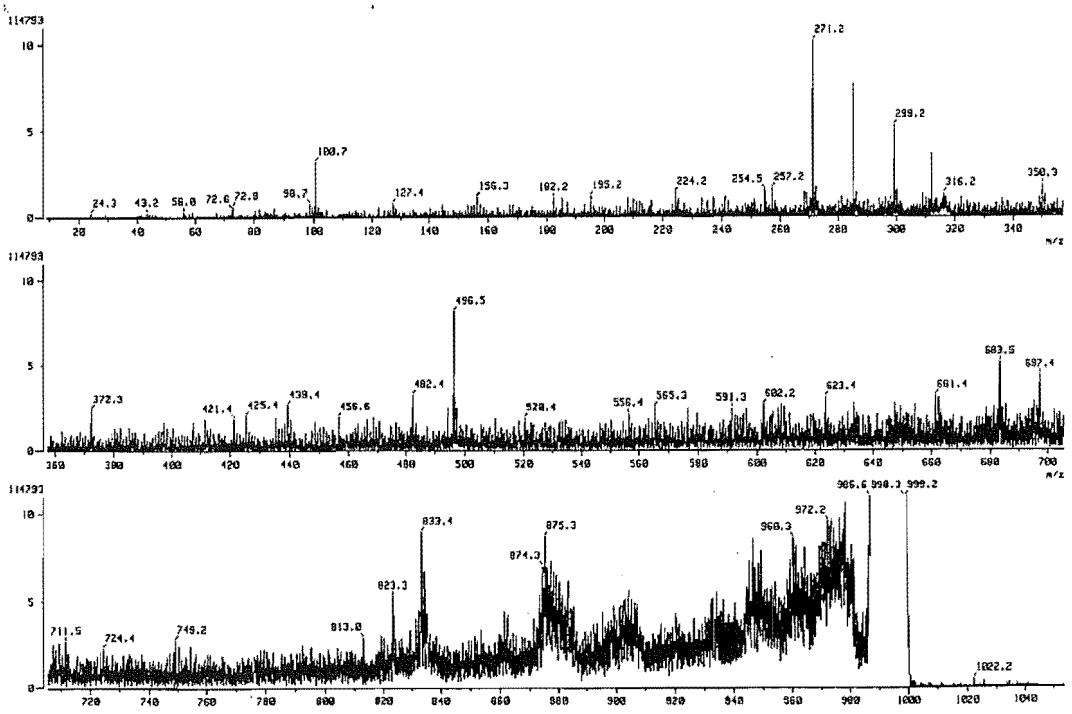


Fig. 6.3.5 Link scan at *m/z* 994

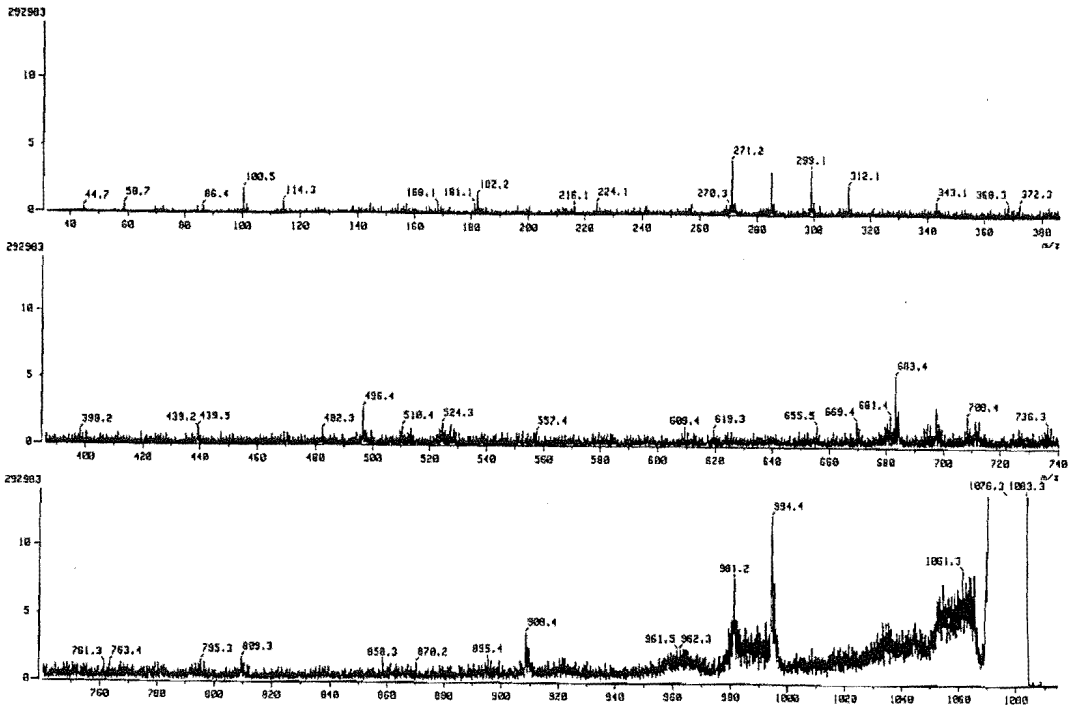


Fig. 6.3.6 Link scan at *m/z* 1079

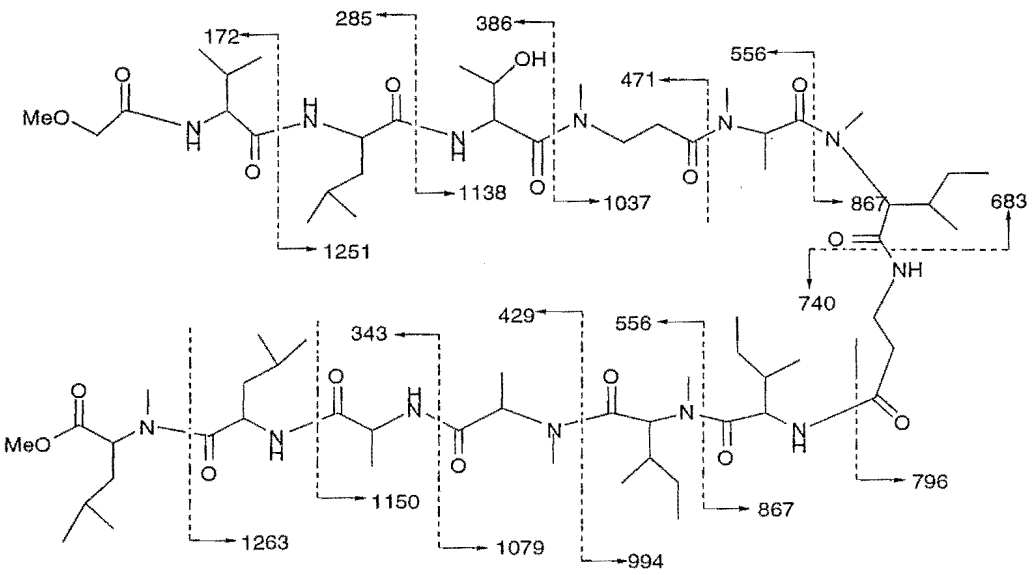
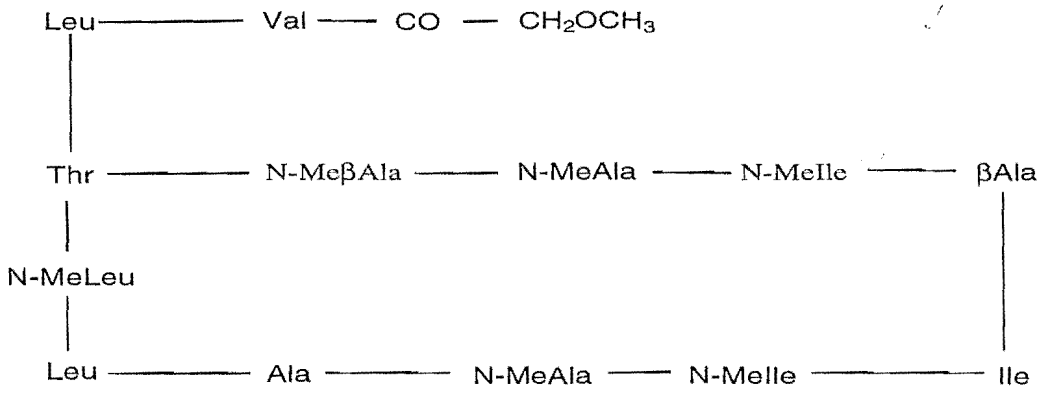


Fig. 6.3.7 FABMS/MS fragmentation from both N- and C-terminus of the ring-opened theonellapeptolide IIIc (**57**)



Sequence of the theonellapeptolide IIIc (**56**)

N.B. Ala and βAla, MeAla and MeβAla, Leu, Ile and *allo*-Ile, and MeLeu, MeIle and MeaIle are isobaric and their positions are exchangeable.

As for IIIa, the spin systems of IIIc were identified by COSY, TOCSY, HSQC and HMBC techniques. The HMBC correlations useful for sequencing included: H2 (methoxyl acetate) and NH (Val)/CO (methoxyl acetate), H2 (Val) and NH (Leu 1)/CO (Val), H2 (Thr) and N-CH_3 (MeβAla)/CO (Thr), N-CH_3 (MeAla 1)/CO (MeβAla), N-CH_3 (MeIle 1)/CO (MeAla), NH (Ile)/CO (βAla), H2 (Ile) and N-CH_3 (MeIle 2)/CO (Ile), H2 (MeIle 2) and N-CH_3 (MeAla 2)/CO (MeIle 2), H2 (MeAla 2) and NH (Ala)/CO (MeAla 2), H2 (Ala)and NH (Leu)/CO (Ala), N-CH_3 (MeLeu)/CO (Leu), N-CH_3 (MeLeu)/C2 (MeLeu), and H2 and H3 (MeLeu)/CO (MeLeu). Due to heavy overlapping of resonances in the ^1H and ^{13}C spectra and missing connectivities, such as those between leucine (1) and threonine, between N-methyl isoleucine (1) and β-alanine, and between N-methyl leucine and threonine, the structure of IIIc can not be unambiguously assigned from the NMR spectra. The sequence of the peptolide was deduced by cross reference to MS/MS analysis and the sequence of the other theonellapeptolides.

Theonellapeptolide IIIc was subjected to acid hydrolysis to give the free amino acids, which were derivatised with Nα(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA) and 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl isothiocyanate (GITC), and analysed by LC/MS. The configuration of each individual amino acid was determined in the same way as in Section 6.2. This peptolide gave a mixture of D-Val, L-Leu, L-Thr, L-N-MeIle, D-Ile, L-N-MeAla, L-Ala, D-Leu and D-N-MeLeu, as well as the non-chiral βAla and MeβAla. A structure showing the stereochemistry of theonellapeptolide IIIc was given in Fig. 6.3.8.

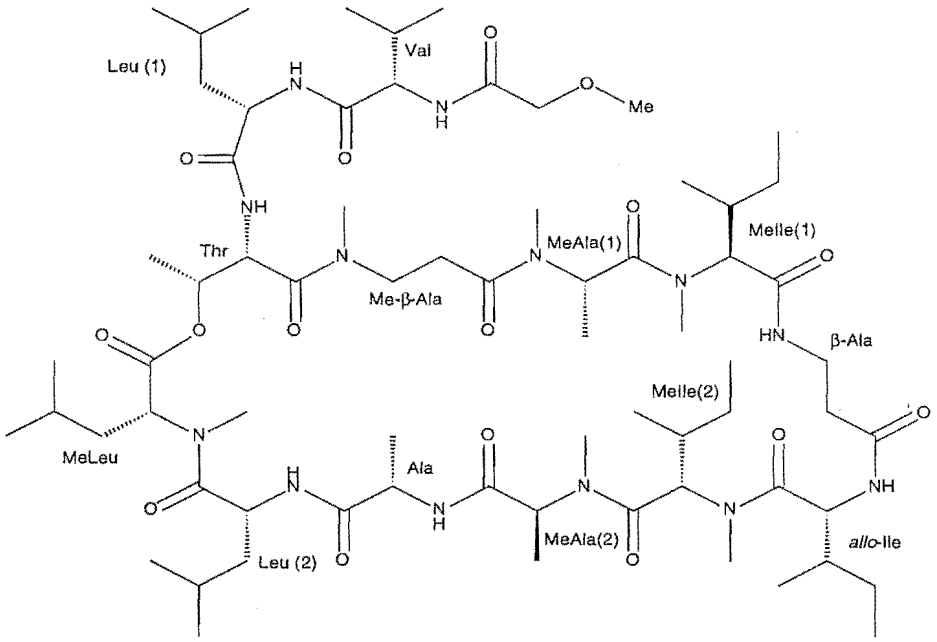
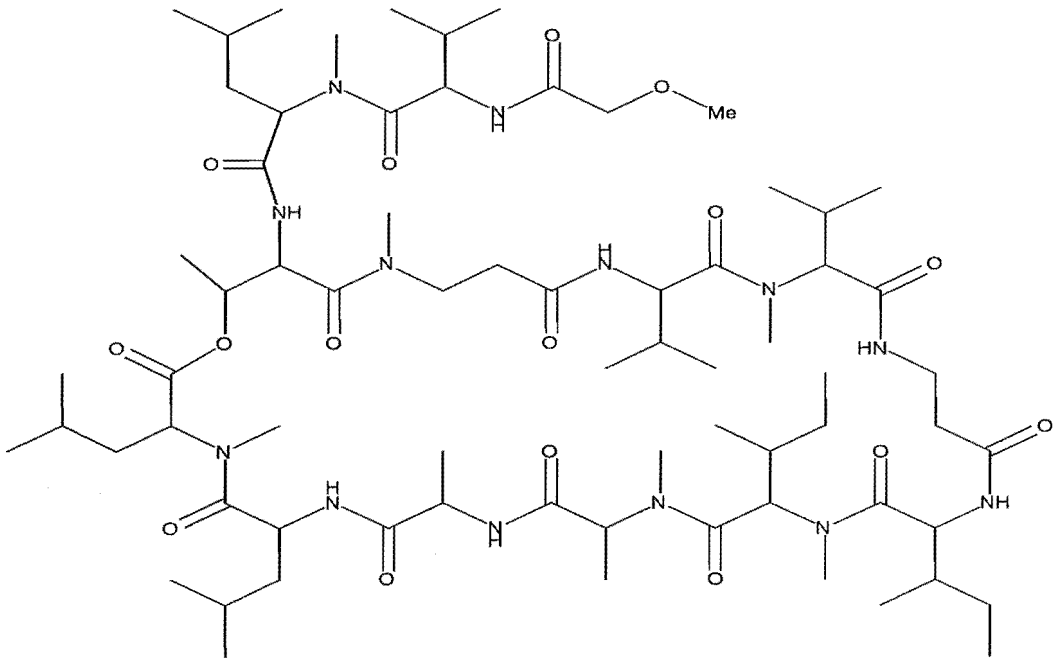


Fig. 6.3.8 Stereochemistry of theonellapeptolide IIIc (**56**)

6.4 Theonellapeptolide IIId

Theonellapeptolide IIId (**58**) gave a pseudo-molecular ion peaks at *m/z* 1404.9 $(M+H)^+$, 1426.9 $(M+Na)^+$ and 1442.9 $(M+K)^+$ in the FABMS spectrum. The molecular formula, $C_{70}H_{125}N_{13}O_{16}Cs$, of **58** was determined by the HRFABMS (*m/z* 1536.8409, M+Cs, Δ -1.2 mmu). The molecular weight of IIId was 14 amu less than that of IIIe. This implied the difference of a methyl group between the two peptolides. There were many similarities in their ^1H NMR spectra. For example, they both exhibited seven amide proton and seven N- or O-methyl resonances; an isolated methylene signal appeared around δ 3.86; most of the α-protons of the α-amino acids appared between δ 4.90 and 5.47. However, the ^1H NMR spectrum of IIId (Fig. 6.4.1) was more complicated than that of IIIe due to the existence of conformers in solution.



Theonellapeptolide IIId (**58**)

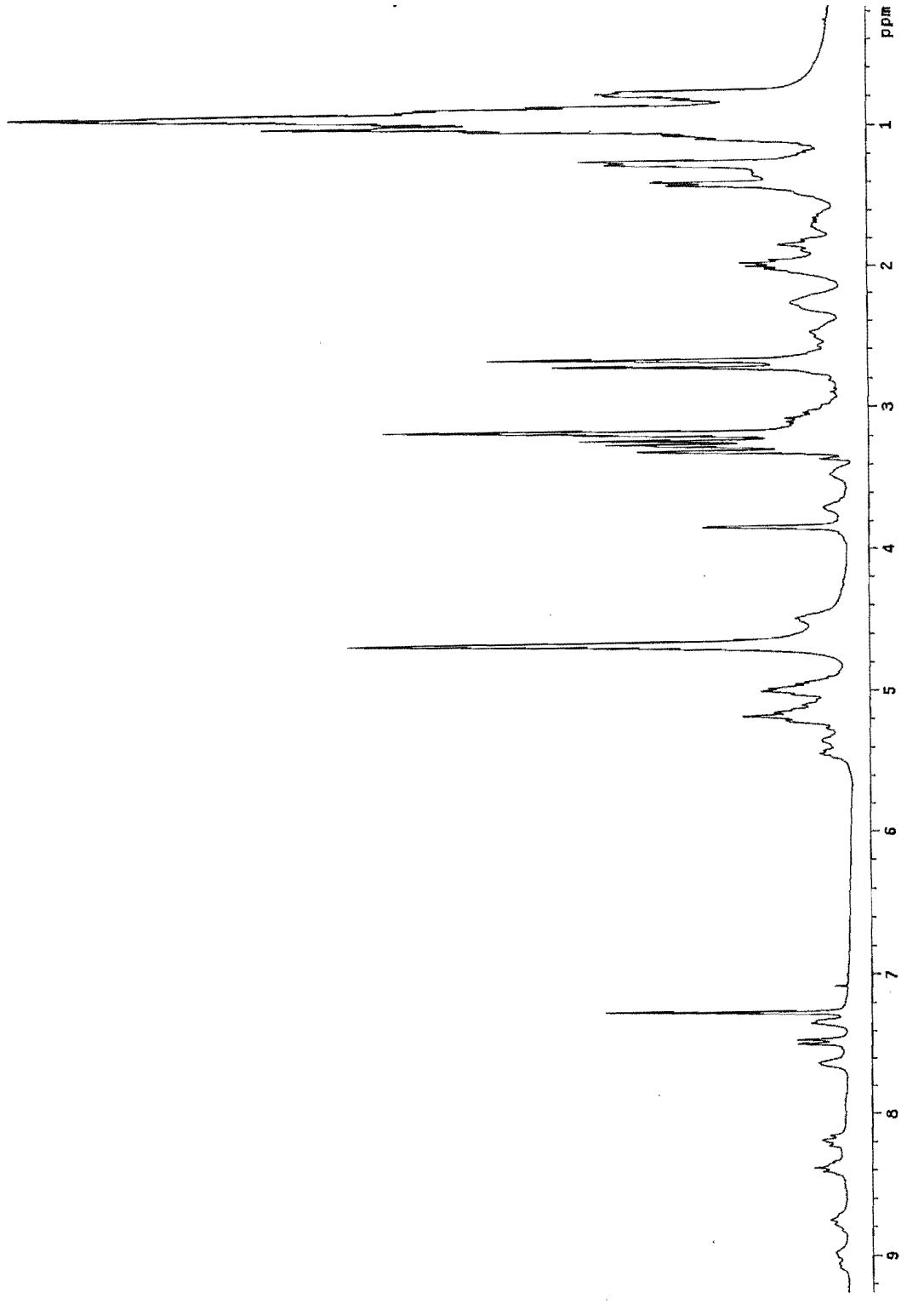


Fig. 6.4.1 ^{1}H NMR spectrum of theonellapeptolide IIId (**58**)

GC/MS analysis of the TFA, butyl ester derivatives of the hydrolysate from IIId resulted in identification of 13 amino acids. This peptolide comprised of threonine, valine (x 2), N-methyl leucine (x 2), N-methyl-β-alanine, β-alanine, isoleucine, N-methyl isoleucine, alanine, N-methyl alanine, leucine and a new amino acid. A GC peak appeared between that for N-methyl leucine and alanine which was unique to this compound. An ion at *m/z* 283 in the mass spectrum of this amino acid derivative suggested either a C6 amino acid or an N-methyl C5 amino acid. An ion at *m/z* 110, characteristic of an N-methyl amino acid, suggested that it was the latter. A relatively intense *m/z* 140 peak, corresponding to loss of an isopropyl group from ion I (see Fig. 4.3.1), strongly implied that the identity of this new amino acid was an N-methyl valine. This assignment was susequently confirmed by NMR analysis.

The sequence of theonellapeptolide IIId was deduced based on two series of ions, 1265, 1138, 1037, 853, 740, 669 (weak), 556, 429 and 344, and 1277, 1164, 1093, 1008, 697, 584 (weak), 400 and 299 observed in the FABMS and daughter ion spectra (Fig. 6.4.2 and 6.4.3) of the ring-opened peptide (**59**). Fragment **59a** belongs to the C-terminal fragmentation ions (Y_i), while fragment **59b** represents N-terminal cleavage (B_i). FAB MS/MS afforded more sequence information for the linear peptide.

The sequences of the tripeptide at the N-terminus and the tetrapeptide at the C-terminus were confirmed by comparison of the product ion spectra of *m/z* 400 and 429 with those of theonellapeptolide IIIe. Ions at *m/z* 952, 853 and 740 found in the spectrum of the link scan at *m/z* 1037 (Fig. 6.4.4) afforded clear evidence for the sequence MeβAla-Val-MeVal (or Ile or Leu). Similarly, the ion at *m/z* 740 yielded daughter ions at *m/z*

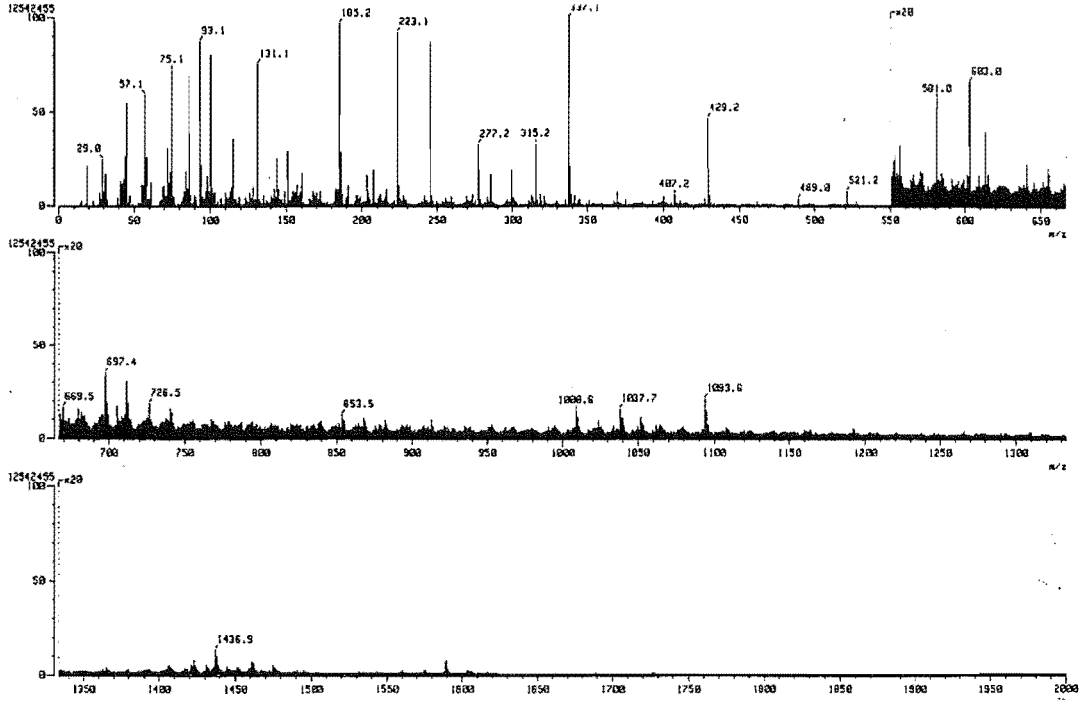


Fig. 6.4.2 FABMS spectrum of compound **59**

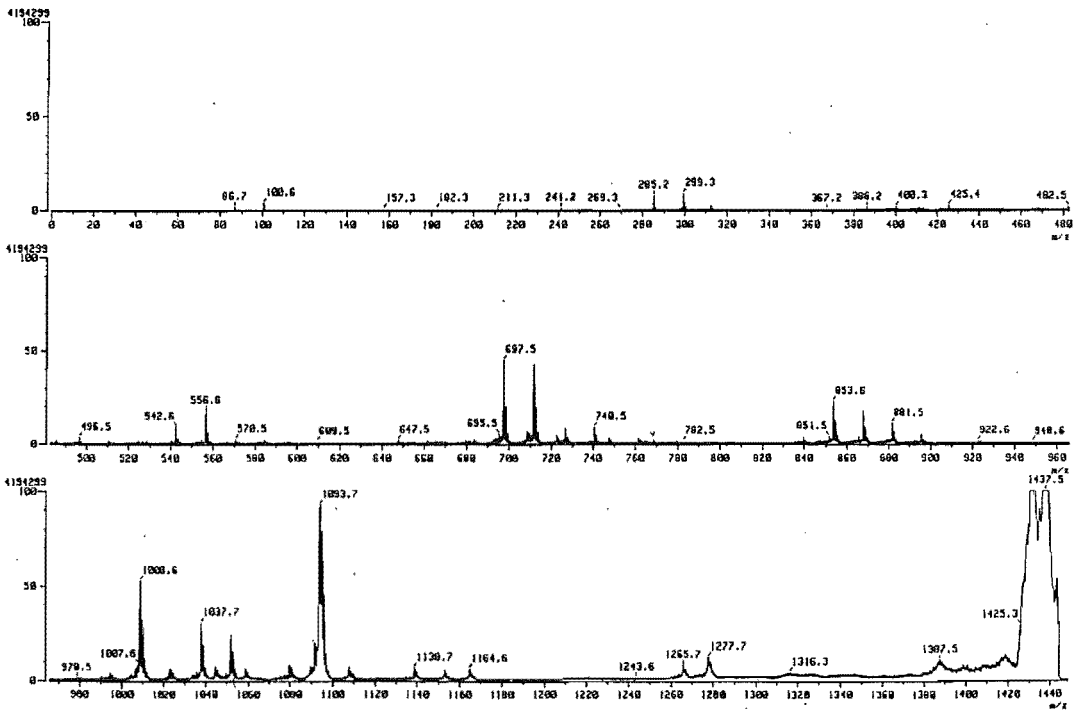
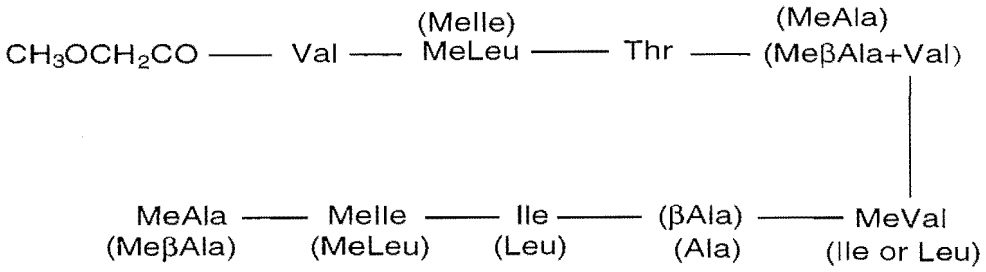
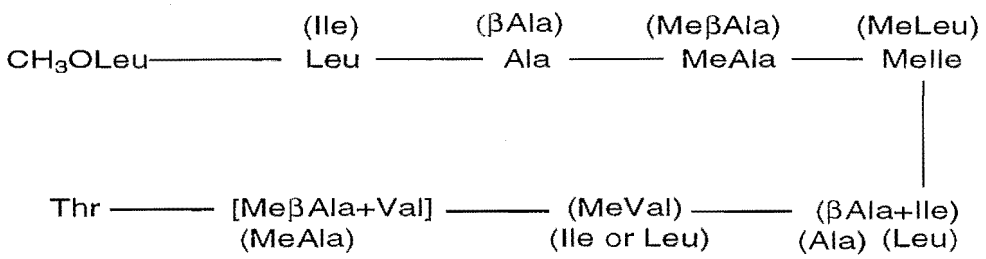


Fig. 6.4.3 Link scan at *m/z* 1437



Peptide fragment **59a**



Peptide fragment **59b**

669 and 556 (Fig. 6.4.5), which was indicative of the seventh and the eighth residues in the sequence: β-Ala-Ile. The mass difference between the fifth and the seventh residue was 113 amu. This can be accounted for by an N-methyl valine unit. The detailed sequencing was depicted in Fig. 6.4.6.

The isobaric amino acids were distinguished by analysis of the 1D and 2D NMR data. The intact amino acid sequence of theonellapeptolide IIId was elucidated primarily by interpretation of the HMBC spectrum. Amide protons at δ 8.99, 8.75, 8.39, 8.17, 7.62, 7.48 and 7.39 were assigned to the spin systems of Thr, Ile, Leu, Val, β-Ala, Val and Ala residues respectively, based on the TOCSY spectrum. While the spin systems related to the methyl amino acids were recognised by either observation of the cross peaks in TOCSY spectrum or by analogy with corresponding

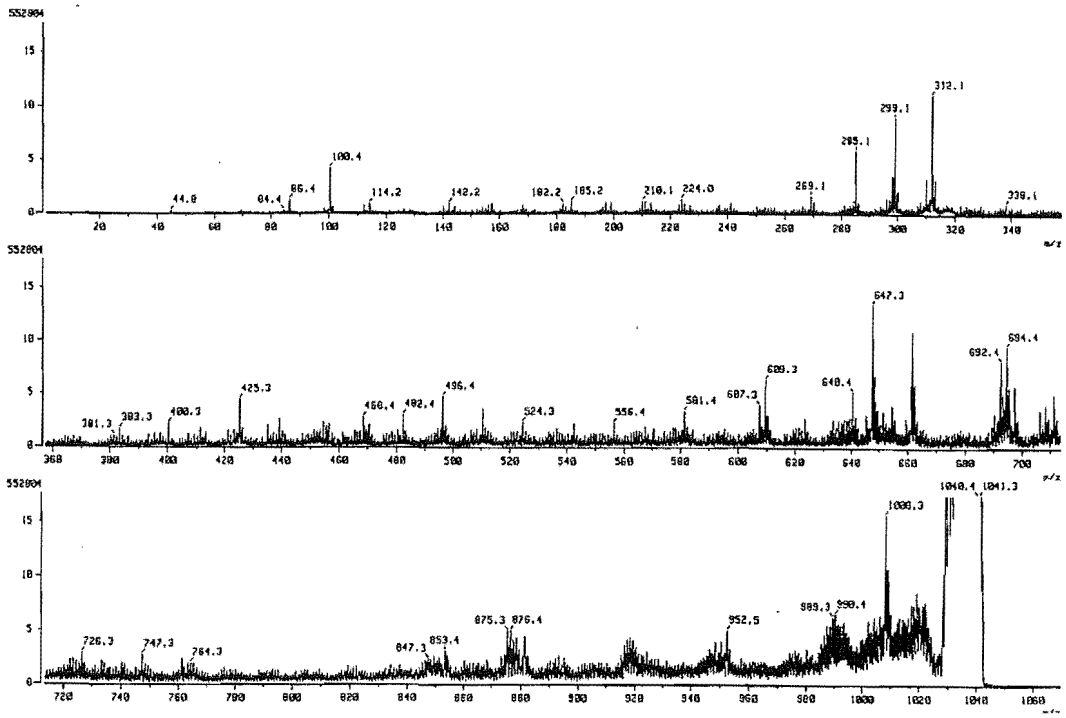


Fig. 6.4.4 Link scan at *m/z* 1037

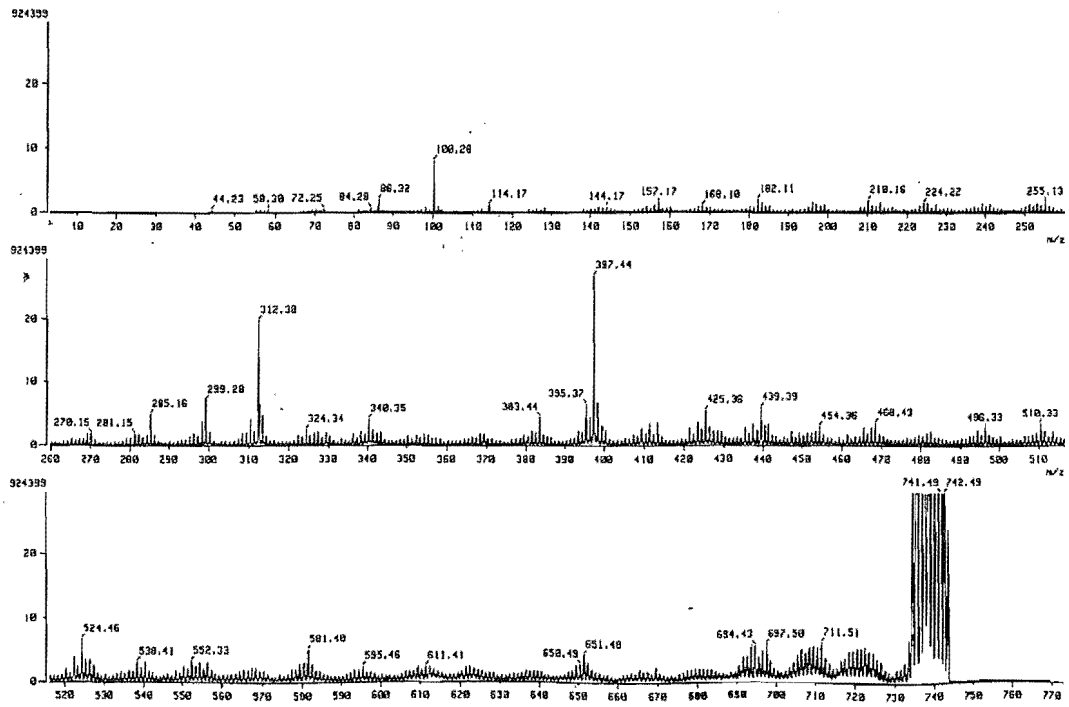


Fig. 6.4.5 Link scan at *m/z* 740

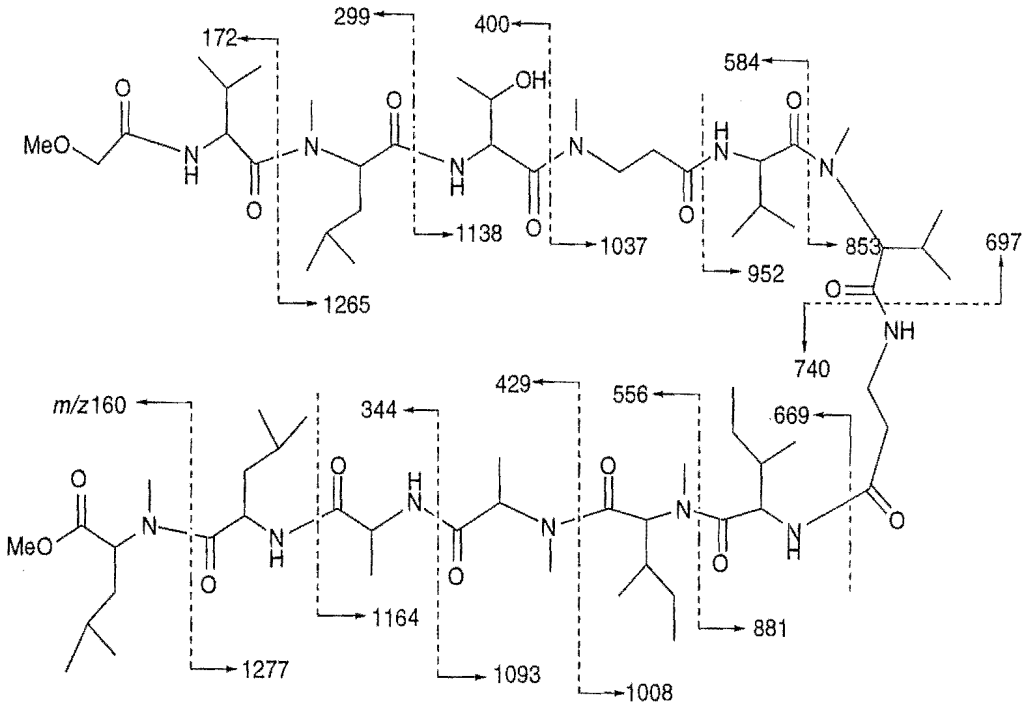


Fig. 6.4.6 FAB MS-MS fragmentation of protonated **59**

spin systems of IIIe. The N-methyl valine residue was defined by subtracting the resonances of N-methyl alanine and N-methyl isoleucine from a TOCSY trace at δ 5.22. The TOCSY spectra of these three residues overlapped due to the close chemical shifts (δ 5.22, 5.18 and 5.18 for MeVal, MeAla and MeIle) of the α-hydrogens (H2). Correlations from two methyl groups at δ 1.13 and 1.06, and from an N-methyl group to the same α-carbon at δ 61 in the HMBC spectrum confirmed the residue.

Inter-residue cross peaks observed in the HMBC experiments were as follows: NH (Val 1)/CO (methoxyl acetate), NH (Val 1)/C2 (Val 1), NCH_3 (MeLeu 1)/CO (Val 1), NCH_3 (MeLeu 1)/C2 (MeLeu 1), H2 (Thr)/CO (MeLeu 1), H2 (Thr) and NCH_3 (MeβAla)/CO (Thr), NH (Val 2)/CO (MeβAla), NCH_3 (MeVal)/CO (Val 2), NH (βAla)/CO (MeVal), H2

(βAla) and H2 (Ile)/CO (βAla), H2 (Ile) and NCH_3 (MeIle)/CO (Ile), NCH_3 (MeIle)/C2 (MeIle), H2 (MeIle) and NCH_3 (MeAla)/CO (MeIle), NCH_3 (MeAla)/C2 (MeAla), H2 (MeAla) and NH (Ala)/CO (MeAla), H2 (Ala) and NH (Leu)/CO (Ala), H2 (Leu) and NCH_3 (MeLeu 2)/CO (Leu), NCH_3 (MeLeu 2) and H3 (MeLeu 2)/C2 (MeLeu 2), and H3 (MeLeu 2) and H3 (Thr)/CO (MeLeu 2). These data led to the sequence of IIId as shown in Fig.6.4.7, which was consistent with the result of the MS/MS analysis.

The absolute stereochemistry of the amino acid residues was determined by HPLC/MS analysis of the derivatised acid hydrolysate , and comparing with those of the standard amino acid derivatives as described in Section 5.6 (also see Experimental). Like the other theonellapeptolides, IIId contains both L- and D-amino acids, which include L-Val, D-MeLeu, L-Thr, D-Val, L-MeVal, D-Ile, L-MeIle, L-MeAla, L-Ala and D-Leu. The stereochemistry of IIId was shown in Fig. 6.4.7.

Theonellapeptolide IIId is cytotoxic against P388 leukemia cells at IC_{50} of 8.8 μg/mL.

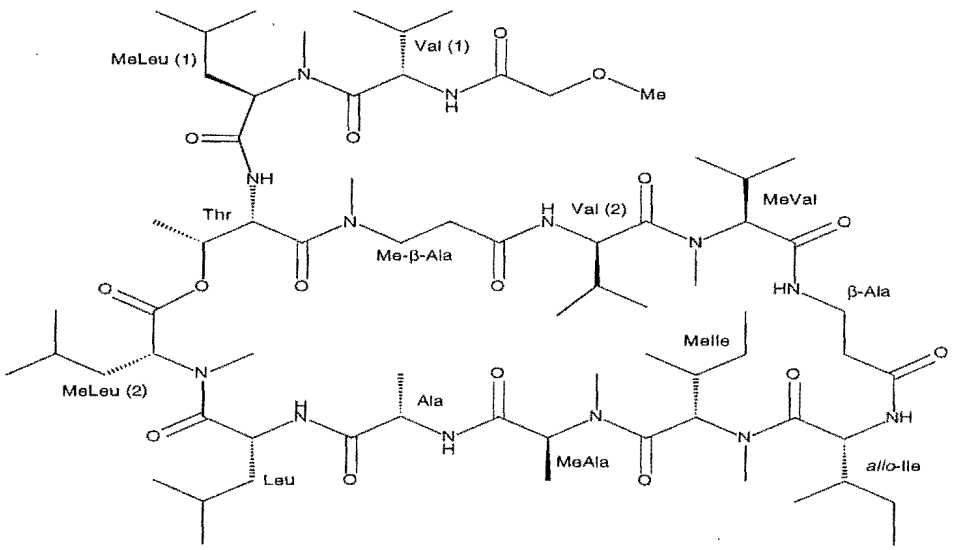


Fig. 6.4.7 Stereochemisry of theonellapeptolide IIId (**58**)

6.5 Other Cytotoxic Peptides from *Lamellomorpha strongylata*

Theonellapeptolide IIIf was the least polar compound in the peptide fraction. This compound had a much longer retention time than IIIe on RPHPLC. FABMS of IIIf exhibited an $(M+H)^+$ ion at 1432.9, accompanied by an $(M+Na)^+$ ion at 1454.8. The ^1H NMR of this compound showed only six amide proton signals between δ 6.8-8.8, and several N- or O-methyl groups between δ 2.6 and 3.3 (Fig. 6.5.1, in $CDCl_3$). The spectral pattern was very much like that of IIIe, and probably only varied in the ratio of the amide protons and N-methyl groups.

The ^1H NMR spectrum of theonellapeptolide IIIg (Fig. 6.5.2) showed at least six broad resonances between δ 7.5 and 9.1, which were attributed to the amide protons, while the sharper signals between δ 2.6-3.3 were probably due to the presence of the N- or O-methyls. In addition, a singlet around δ 3.85 was assigned to the methylene of a methoxyl acetate moiety. Thus, this compound can be considered as having a similar structural skeleton to the other theonellapeptolide components.

IIIh was the least abundant component in this group of peptide compounds (~2 mg only was obtained). The ^1H NMR spectrum of IIIh (Fig. 6.5.3) was very similar to that of IIId, except for an amide resonance at δ 8.76. This was at a higher field than the comparable resonance in IIId. A similar structure to IIId was suggested for this minor peptide.

Theonellapeptolides IIIg and IIIh exhibited cytotoxicity against P388 cell line with IC_{50} of 5 and 11 μg/mL respectively. IIIf was non-cytotoxic.

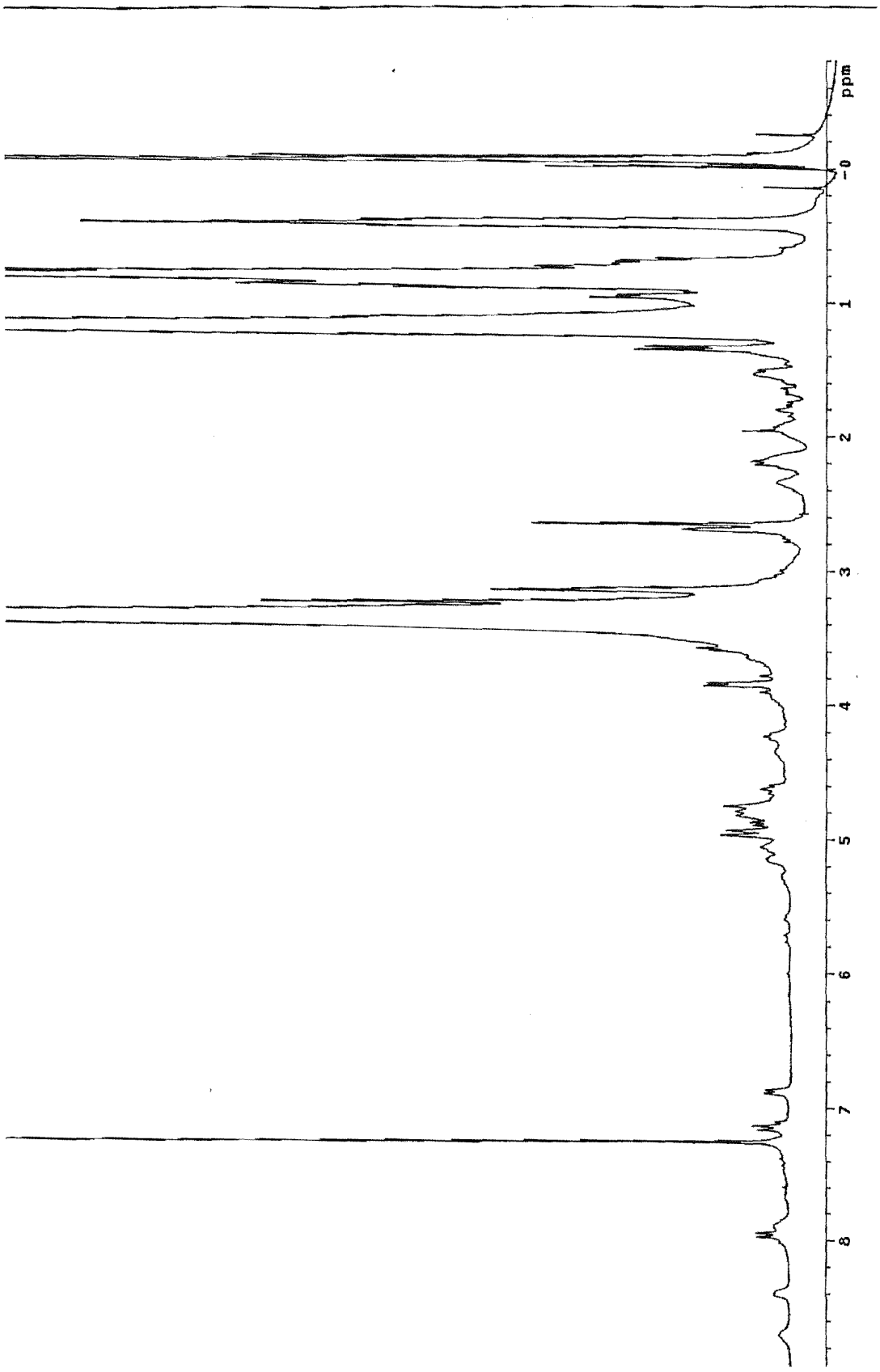


Fig. 6.5.1 ^{1}H NMR spectrum of IIIf

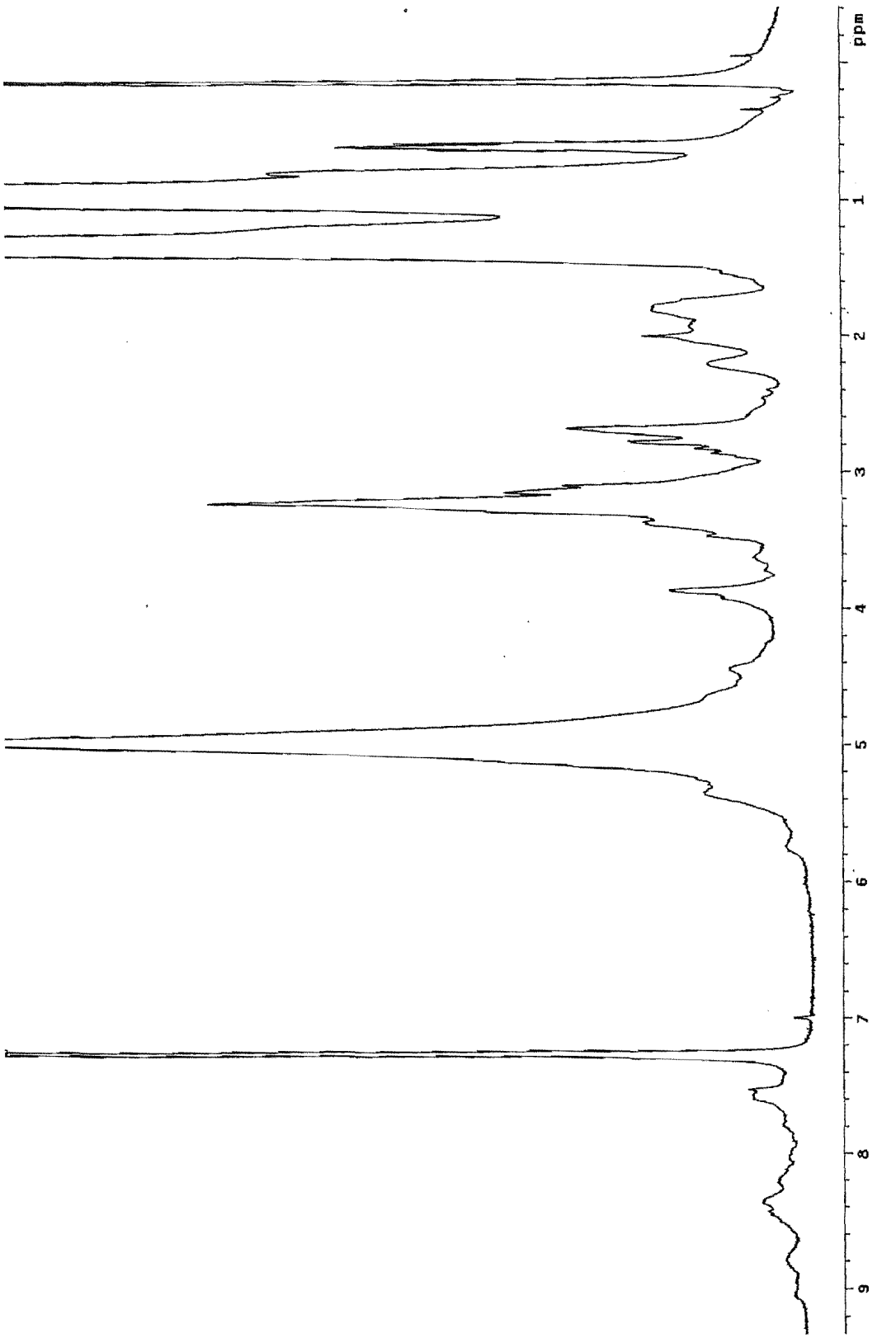


Fig. 6.5.2 ^{1}H NMR spectrum of IIIg

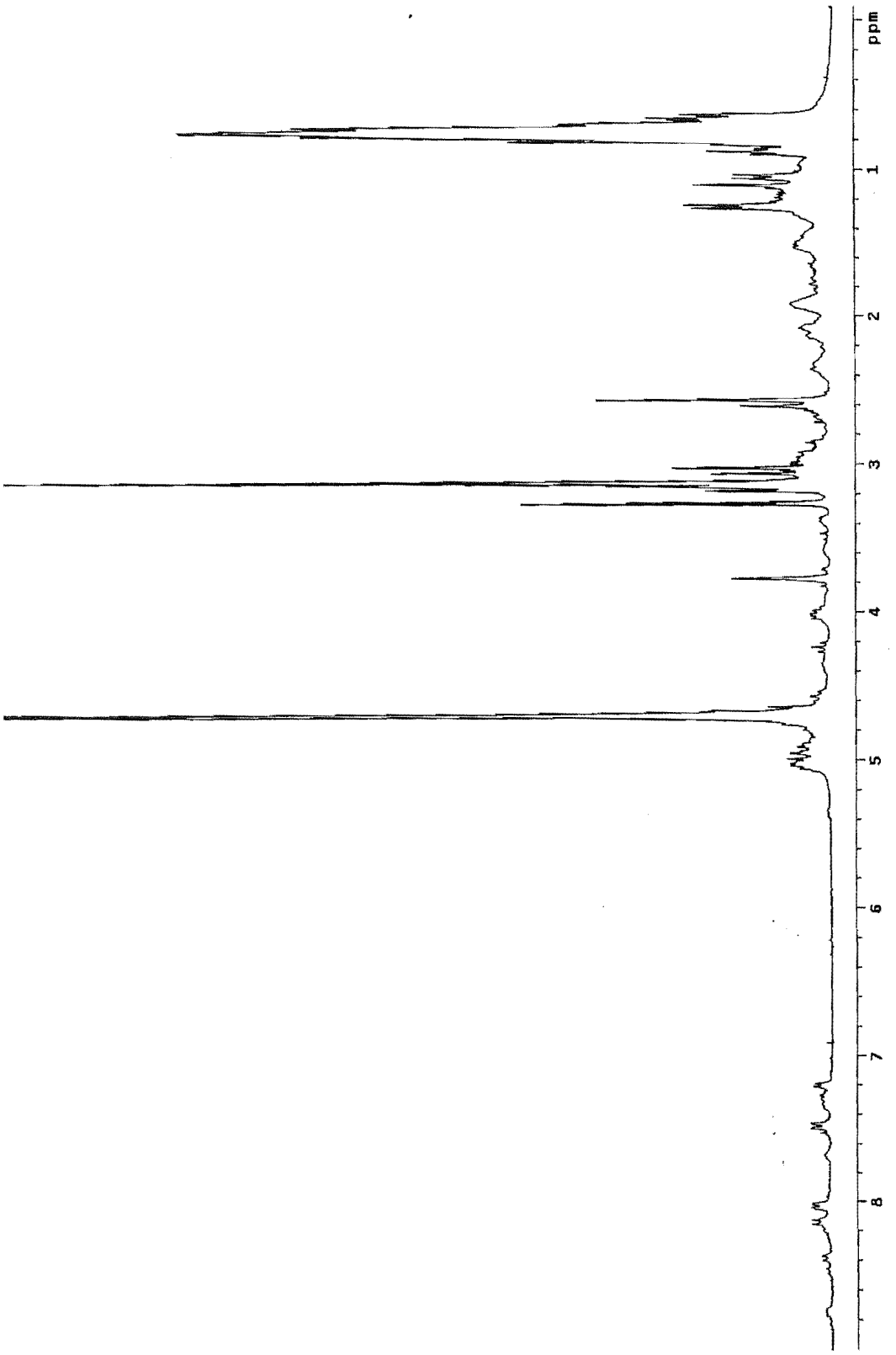


Fig. 6.5.3 ^{1}H NMR spectrum of IIIh

6.6 Structural Variation of the Peptolides

The structural variation of the theonellapeptolides IIIa-e is shown , in summary, in Fig.6.6.1

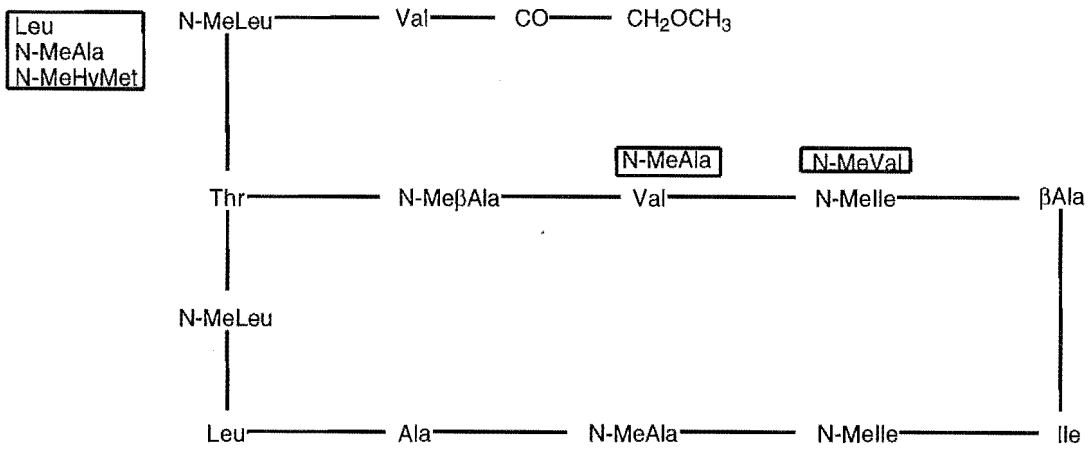


Fig. 6.6.1 Structural comparison of theonellapeptolides IIIa-e

The base structure used was that of IIIe, the major peptolide. The residues in boxes indicate the variations which generated the different theonellapeptolides in the series. The structural changes occurred only at second, fifth or sixth residues with the second residue being that most commonly showing variation. In IIIa the second residue became N-methyl-hydroxymethionine, while in IIIb N-methylalanine occupies the second position. In IIIc, two amino acids are different from those in IIIe. The second residue of IIIe has been replaced by Leu, while valine 2 at the fifth position was replaced by N-methyl alanine. Finally, when the N-MeIle at the sixth position of IIIe was replaced by N-MeVal, the peptolide became IIId.

Theonellapeptolides IIIb, c, d, e all exhibited mild cytotoxicity against P388 cell line, but IIIa was not cytotoxic. This result implies that the nature of the second residue is very important to maintain the bioactivity of this class of compound.

Chapter Seven

Experimental

7.1 General Methods

7.1.1 Nuclear Magnetic Resonance

^{1}H NMR spectra were recorded on a Varian UNITY 300 NMR spectrometer at 23°C, operating at 300 MHz. ^{13}C and DEPT NMR spectra were recorded on a Varian XL 300 NMR spectrometer, at 23°C, operating at 75 MHz. 1D- and 2D-TOCSY, COSY, ECOSY and ROESY, and the reverse-detected HSMQC (or HSQC) and HMBC experiments were recorded either on a Varian UNITY 300 NMR spectrometer, or a Varian INOVA 500 NMR spectrometer. The UNITY 300 instrument was fitted with a Nalorac Z. spec MID300 3mm Indirect Detection Probe. Chemical shifts in this thesis are described in parts per million (ppm) on the δ scale, and were referenced to the appropriate solvent peaks (except otherwise specified): $CDCl_3$ or $CDCl_3/CD_3OD$ referenced to $\mathbf{CH}Cl_3$ at δ_H 7.25 (^{1}H) and $\mathbf{C}HCl_3$ at δ_C 77.0 (^{13}C); CD_3OD or CD_3OH referenced to $CD_3O\mathbf{H}$ at δ_H 4.70 (^{1}H) and $\mathbf{C}D_3OD$ at δ_C 49.3 (^{13}C); D_2O referenced to $\mathbf{H}OD$ at δ_H 4.70 (^{1}H); C_6D_6/CD_3OD referenced to $C_6\mathbf{H}D_5$ at δ_H 7.27 (^{1}H) and $\mathbf{C}_6D_6$ at δ_C 128.4 (^{13}C). ^{1}H NMR spectra were recorded using an aquisition time (AT) of 2 s, except for those spectra run in CD_3OH, where suppression of the CD_3OH solvent peak was run under the following conditions; relaxation delay (D1) =1 s, decoupler power = 4, steady state = 4, decoupler mode = 'ynn'. ^{13}C NMR spectra were recorded using an acquisition time of 0.5 s. Decoupling experiments were run with decoupler power 10 dB. ROESY experiments were run using an AT = 0.46 s, D1 = 0.60 s and a mixing time of 0.30 s. COSY experiments were recorded using an acquisition time of 0.24 s and a relaxation delay of 1 s. 2D-TOCSY spectra were recorded using an acquisition time of 0.50 s, a relaxation delay of 0.60 s and a mixing time

of 80 ms. 1D-TOCSY spectra were recorded using an acquisition time of 2 s, a relaxation delay of 1 s and mixing times as indicated in the discussion. HSMQC experiments were recorded with an acquisition time of 0.2 s, a relaxation delay of 0.4 s and $J_{X\text{-}H}$=130 Hz. HMBC experiments were recorded with an acquisition time of 0.21 s, a relaxation delay of 0.3 s, J = 140 Hz and J_{NXH} = 8.3 Hz, or J_{NXH} = 8.3, 5.0 Hz arrayed mixing times.

7.1.2 Mass Spectrometry

Mass spectra were obtained on a Kratos MS80RFA mass spectrometer, equipped with a Carlo Erba MFC500 GC, operated at 4 kV. FAB was performed with an Ion Tech ZN11FN ion gun using Xe as the reagent gas, operating at 8 kV and 2 mA with either NOBA (*m*-nitrobenzyl alcohol), m-b (magic bullet, 50% thioerythritol/thiothreitol), or glycerol as matrix.

HRFABMS and FABMS/MS were measured on a JEOL SX102A spectrometer (NIDDK, NIH, U.S.A.). NOBA, magic bullet and glycerol were used as matrix.

7.1.3 GC, GC/MS and HPLC/MS

Gas chromatography (GC) was accomplished on a Varian 3700 gas chromatograph equipped with a Hewlett Packard 3390A Integrator. GC/MS spectra were measured on a Kratos MS80RFA mass spectrometer equipped with a Carlo Erba MFC500 GC. LC/MS was carried out on an HP1100 Integrated LCMS System with a C18 Stable Bond Zorbax (150 x 4.6 mm) column.

7.1.4 High Performance Liquid Chromatography

Analytical HPLC was performed on a Philips PU4100 Liquid Chromatograph equipped with a Philips PU4120 Diode Array Detector interfaced to a PC General 486 computer running Philips PU6003 Diode Array Detector System Software (V3.0) and a Hewlett Packard 7475A Plotter. Semipreparative HPLC was performed on a Shimadzu LC-4A Liquid Chromatograph equipped with SPD-2AS Spectrophotometric Detector. H_2O was purified using a MilliQ deionising system. MeOH (J.T. Baker, HPLC reagent grade) and acetonitrile (BDH, HiPerSolv) solvents were used undistilled. All other solvents were of technical grade and were purified and re-distilled before use. Hexane was prepared from petroleum ether (10 L, b.pt. 60-70°C), nitrated with a mixture of *c.* HNO_3 (1 L) and *c.* H_2SO_4 (1 L) for two days. After this time, the nitrating mixture was removed and the organic layer washed with H_2O (20 L), and dried over $CaCl_2$ overnight. The resulting solution was filtered through activated alumina (Laporte, 100 mesh) and distilled. All solvents (with the exception of MilliQ water) were filtered through a Millipore filter apparatus, equipped with a Rainin Nylon-66 filter (0.45 μm) prior to use. All samples were filtered through a syringe filter (Alltech, 0.45 μm) prior to injection.

The analytical reverse phase C18 column used were Microsorb-MV, 86200-C5, 5 μm, 100 Å, 250 x 4.6 mm, and 3 μm, 100 Å, 100 x 4.6 mm (Rainin Instrument Company Inc, USA). The analytical Diol column used was a LiChrospher® 5 μm, 100 Å, 250 x 4.6 mm (E. Merck, Darmstadt, Federal Republic of Germany). The analytical CN column used was a ZORBAX® CN, 5 μm, 250 x 4.6 mm (Dupont Instruments). The

semipreparative reverse phase 18 column used was a Microsorb-MV, 80 299-C5, 5 μm, 100 Å, 250 x 10 mm (Rainin Instrument Company Inc, USA).

Medium pressure column chromatography was performed on a Lobar chromatography setup, with a variable speed electric pump, to achieve a typical flow rate of 10 mL/min. The Diol lobar column used was a Lichroprep® DIOL, 40-63 μm, 310 x 25 mm (E. Merck, Darmstadt, Federal Republic of Germany).

7.1.5 Column Chromatography

All column chromatography was performed on various sizes of glass columns. "Flash" columns were run under N_2 gas (oxygen free) pressure (0.5 kPa). Solvents used were all commercial grade, distilled once in a glass distillation apparatus, except methanol (MeOH), which was distilled twice.

Reverse-phase C18 packing used for preparative and flash column chromatography was prepared from silica gel (Davisil, 35-70 μm) using a solution of octadecyltrichlorosilane (2 % *v/v*, EGA-Chemie) in carbon tetrachloride.[162] The silica gel was stirred gently with a teflon-coated magnetic follower overnight before washing with carbon tetrachloride (x 2). Unreacted chloro groups were removed by washing with dry MeOH. A carbon tetrachloride solution of trimethylchlorosilane (2 % *v/v*, Aldrich) was used to end-cap any unreacted hydroxyl groups. The reverse-phase material was finally washed with dichloromethane (CH_2Cl_2) (x 2) and MeOH (x 2) to remove the reagents. The C18 reverse-phase material was extensively washed with MeOH immediately before use.

Small scale preparative C18 reverse-phase column chromatography was performed using Alltech disposable cartridges (200 mg). Bakerbond Diol (Bb)(40 μm APD) or Merck Diol (Mk)(40 μm APD), and Alltech Silica gel (35-70 μm) were used for normal phase chromatography.

Pre-swollen Sephadex LH-20 (Pharmacia Biotic AB) was used for gel permeation chromatography. Sephadex G-25 (SIGMA) was used for gel filtration chromatography.

Celite filtration and coating was performed using diatomaceous earth (Celite Corporation).

7.1.6 Thin Layer Chromatography

The silica analytical TLC described in this thesis was performed on Merck silica gel 60 F_{254} aluminium-backed sheets, 0.2 mm in thickness. Diol analytical TLC was performed on Merck F_{254} glass-backed plates of 0.2 mm thickness. Whatman MKC ^{18}F TLC plates of 0.2 mm thickness were used for analytical C18 TLC. Amino analytical TLC was performed on Merck F_{254} glass-backed plates of 0.2 mm thickness. All TLC plates were visualised initially under short-wave (λ 254 nm) light, or, subsequently, with a phosphomolybdic acid in ethanol spray (10 % phosphomolybdic acid in ethanol *w/v*), except for analysis of theonellapeptolides. Those TLC plates were visualised with I_2.

7.1.7 Dry Solvents

Dichloromethane was dried initially over calcium chloride then distilled from calcium hydride. Dry tetrahydrofuran was achieved by distillation over calcium hydride. Dry methanol was distilled from magnesium metal and iodine. Dry isopropyl alcohol was obtained by refluxing with

calcium oxide, and then distillation. n-Butanol was treated with anhydrous calcium sulphate, and fractionated after filtration from the desiccant. All dry solvents were distilled immediately prior to use.

7.1.8 Other Experiments

MPs were determined on an Electrothermal melting point apparatus and are uncorrected. Optical rotations were measured on a JASCO J-20 recording spectropolarimeter. IR spectra were recorded on a Shimadzu FTIR-8201PC infrared spectrophotometer. High speed countercurrent chromatography was run on an Ito Multi-Layer Coil Separator-Extractor. The X-ray diffraction data was collected on a Siemens SMART area detector system.

7.2 Work Related to Chapter 2

7.2.1 Chemical Screening of the Organic Extract from *Mycale* sp.

Organic extracts from *Mycale* sp. were dissolved in 50% DCM/MeOH (reserved 200 μL of this solution as the primary standard, and 200 μL aliquot for gel permeation chromatography), and coated onto celite to make a material containing 20 mg extract per 100 mg celite. The C18, Diol, PEI WP, CBA WP and LH-20 cartridges were chosen for testing of this sample. All cartridges, except LH-20, were Bond Elut solid phase extraction columns from Analytical International. The LH-20 column was prepared in the laboratory using the previously described method.[123] The selected cartridges were pre-equilibrated with MeOH (3 mL), followed by the initial eluting solvent (10 mL). The organic extract (10 mg) (either 50 mg coated celite, or 200 μL solution for LH-20) was loaded onto each of the cartridges. The volume of each solvent used in elution was 3 mL, except for the initial solvent (2 mL) for PEI WP and CBA WP, and the three fractions (1.3, 1.5 and 4 mL) from LH-20 column. A total of 14 fractions were collected, including F1 (50% MeOH/H_2O), F2 (MeOH) and F3 (50% MeOH/DCM) from C18, F4 (DCM), F5 (EtOAc) and F6 (MeOH) from Diol, F7 (0.025 M NH_4Ac), F8 (0.5 M HOAc) and F9 (0.5 M sodium citrate) from PEI WP, F10 (0.025 M NH_4Ac) and F11 (0.5 M HOAc) from CBA WP, and F12, F13 and F14 (50% MeOH/DCM) from LH-20. The P388 assay results (IC_{50}, dilution) for these collections were summarised as follow: F1, 230; F2, <97.5; F3, 310; F4, 768, F5, 1373; F6, <97.5; F7, 2758; F8, <97.5; F9, 120; F10, 455; F11, <97.5; F12, 286; F13, <97.5; F14, 115; and standard, <97.5.

All 14 fractions were analysed by RPHPLC using a C2 column. The elution was made in a gradient mode with acetonitrile/water from 20% through to 80%. The eluate was monitored at 225, 254 and 270 nm with a PU 4120 diode array detector. The peak corresponding to pateamine was observed as one of the major peaks in the HPLC traces of F2, F6, F8, F11 and F13. It was also recognisable in the trace of F14.

7.2.2 Chemical Screening of the Organic Extract from *Lamellomorpha strongylata*

The organic extract from *L. strongylata* was analysed by the same chemical screening procedure as used in Section 7.2.1. The cytotoxicity test revealed F1, 9908; F2, <97.5; F3, 106; F4, <97.5; F5, <97.5; F6, <97.5; F7, <97.5; F8, <97.5; F9, 160; F10, <97.5; F11, <97.5; F12, <97.5; F13, <97.5; F14, 574; and standard, <97.5.

All 14 fractions were processed by RPHPLC on an ODS column. Gradient elution was used with an acetonitrile/water system (containing 0.05% TFA), starting a 40% and ending up with 100% acetonitrile (monitored at 214, 268 and 344 nm). The peaks with retention time 320 and 580 secs, which corresponded to swinholide and calyculin compounds respectively, were readily recognised in the HPLC traces of all the active fractions (IC_{50}<97.5).

7.2.3 Study of Anti-HIV components from *Chondropsis kirkii*

Collection The sponge was collected by SCUBA at Lighthouse Reef Point, Moeraki, South Iland, New Zealand, in 1993. *Chondropsis kirkii* belongs to the family *Desmacidonidae*, order *Poecilosclerida.*

Extraction Frozen sponge (380 g) was crushed and soaked in MeOH (300 mL) overnight. After blending, the sample was filtered through a sintered funnel. The residue was extracted with MeOH (2x150 mL), 50 % MeOH/DCM (150 mL), and DCM (150 mL) for one hour for each extraction. These extracts were combined. After removing the organic solvents from the extract, an equal volume of absolute ethanol was added to form a 50% aqueous alcohol solution. Most of the biopolymers, such as proteins and polysaccharides, were precipitated, and removed by filtration. The aqueous portion (after removal of the EtOH under vacuum) was partitioned against EtOAc (3x100 mL). The combined EtOAc extract was the organic portion (558 mg). The aqueous solution was freeze-dried to afford the aqueous extract (2.12 g).

Chemical screening of the Aqueous Extract A C4 WP cartridge and a Sephadex G-25 cartridge were chosen for analysis of the extract. The Sephadex G-25 column was prepared by putting the H_2O-soaked Sephadex G-25 gel (1 g) into a 6 mL empty cartridge. Both columns were pre-equilibrated with H_2O, and loaded with the aqueous sample. The C4 WP column was eluted with H_2O, 33% MeOH/H_2O, 66% MeOH/H_2O and MeOH. The Sephadex G-25 column was eluted with H_2O. Four fractions from each column were collected, and tested for both anti-HIV and cytotoxic activities. The former was evaluated by the inhibition %, while the latter was judged by the percentage of killed cells. The anti-HIV and cytotoxic activities (in brackets) were recorded as (%): F1, 136 (36); F2, 5 (0); F3, 1 (56); F4, 0 (14) from C4; F5, 106 (0); F6, 86 (0); F7, 0 (14); F8, 0 (9) from Sephadex G-25.

Chemical screening of the Organic Extract C18, Diol and PEI WP cartridges were selected for processing of the extract. The solvents and operations were the same as those used in Section 7.2.1. The assay

results were summarised as (the fractions were numbered in elution order from individual column and in a series of C18, Diol and PEI) HIV inhibition% and cytotoxicity% (in brackets): F1, 0 (31); F2, 0 (99); F3, 54 (31); F4, 51 (34), F5, 0 (100); F6, 0 (100); F7, 0 (100); F8 0 (100); F9, 0 (100).

Sephadex G-25 Chromatography of the Aqueous Extract The aqueous extract (50 mg) was dissolved in H_2O (500 μL), and loaded onto the head of a G-25 column, containing Sephadex G-25 (15 g). The solvent (H_2O) was delivered with an LKB 2132 Microperpex Peristaltic pump. 22 Fractions (25 mL/fraction) were collected. They were combined into 9 fractions corresponding to 9 peaks. Bioassay tests afforded the following results (cytotoxicity% was in parenthesis): F1, 103 (13); F2, 73 (11); F3, 0 (28); F4, 0 (15); F5, 0 (17); F6, 0 (16); F7, 0 (20); F8, 0 (99); F9, 0 (15).

A sample of the precipitate (20 mg) from 50% aqueous ethanol was redissolved in H_2O, and analysed in the same way as the supernatant on Sephadex G-25 column. The five of seven resultant fractions (labelled with P1-P5) showed strong anti-HIV activity and low cytotoxicity (in parentheses): P1, 91 (10); P2, 107 (11); P3, 113 (21); P4; 103 (30), P5, 103 (4).

Toluidine Blue Dye Binding Test Toluidine blue (1 mg) was dissolved in H_2O (100 mL) to give a blue-coloured solution (10 μg/mL). A 10% aqueous dextran sulphate was used as positive standard. To toluidine blue (1 mL), was added series of Dextran sulphate standards (10, 20, 30 and 40 μL). To toluidine blue (2 mL), were also added dextran sulphate standards (10, 20, 30, 40, 50, 60 and 70 μL). Samples (10, 20 and 50 μL) from each fraction (before combination) from the

Table 7.2.1 Toluidine blue test results

Sample Name	Concentration (μL/2 mL)	Absorbance (A)	Sample Name	Concentration (μL/2 mL)	Absorbance (A)
Standard (1 mL)	0	0.83742	F8	10	0.81760
	10	0.64651		20	0.78391
	20	0.42670		50	0.79201
	30	0.21564			
	40	0.17943			
Standard (2 mL)	0	0.83240	F9	10	0.80745
	10	0.66655		20	0.75948
	20	0.56210		50	0.77408
	30	0.45123			
	40	0.35185			
	50	0.28566			
	60	0.21336			
	70	0.17464			
P1	10	0.72450	F10	10	0.78522
	20	0.72110		20	0.77020
	50	0.69055		50	0.74821
P2	10	0.74402	F11	10	0.78758
	20	0.68571		20	0.80524
	50	0.59718		50	0.79704
P3	10	0.69980	F12	10	0.77457
	20	0.65649		20	0.77571
	50	0.56662		50	0.72334
P4	10	0.72490	F13	10	0.68822
	20	0.67249		20	0.70253
	50	0.55156		50	0.63510
P5	10	0.70126	F14	10	0.63884
	20	0.68712		20	0.67935
	50	0.59273		50	0.70958
P6	10	0.74568	F15	10	0.63194
	20	0.71267		20	0.68692
	50	0.65625		50	0.71089
P7	10	0.71436	F16	10	0.71570
	20	0.69433		20	0.72198
	50	0.68712		50	0.70078
F1	10	0.78510	F17	10	0.70349
	20	0.79169		20	0.69724
	50	0.77525		50	0.69475
F2	10	0.75233	F18	10	0.68693
	20	0.71169		20	0.66440
	50	0.55945		50	0.68764
F3	10	0.79483	F19	10	0.68958
	20	0.74599		20	0.69745
	50	0.72418		50	0.70332
F4	10	0.76369	F20	10	0.77040
	20	0.74011		20	0.72729
	50	0.79123		50	0.69064
F5	10	0.80962	F21	10	0.70761
	20	0.79553		20	0.71284
	50	0.77437		50	0.70232
F6	10	0.80414	F22	10	0.78301
	20	0.77777		20	0.76959
	50	0.77676		50	0.77655
F7	10	0.78522			
	20	0.74899			
	50	0.79402			

G-25 column on the aqueous extract were each then added to toluidine blue (2 mL). The absorbance of all the sample-dye mixtures were measured on a Hewlett Packard 8452 diode array spectrophotometer at 620 nm. The results are summarised in Table 7.2.1.

Isolation of HIV-inhibiting compounds from *C. kirkii* An initial separation of the anti-HIV components from the organic extract was achieved using C-18 reverse phase flash chromatography, monitoring by bioassay and Diol TLC (1% MeOH/DCM). Further column chromatography on Diol (13 g) of the active fractions, using 30% DCM/petroleum ether through to 70% DCM/petroleum ether, concentrated the activity into three fractions, eluted by 30% DCM/petroleum. The active fractions were loaded onto an LH-20 (40 g) column, which was eluted with 10% MeOH/DCM (10 mL/fraction). ^{1}H NMR spectroscopy indicated that fractions 2 and 3 contained mainly steroidal compounds, and exhibited medium anti-HIV activity. These fractions were combined and further purified on a CN HPLC column (4.6 mm x 25 cm, in normal phase mode) using 5% DCM/hexane as eluent. A total of seven fractions were collected. The bioassay results from the selected fractions are given in Table 2.3.1.

The steroidal fraction (29 mg) was dissolved in pyridine (300 μL), to which an equal amount of acetic anhydride was added. The reaction mixture was shaken, and then left at room temperature overnight. The solution was transferred into a separating funnel, diluted with water (6 mL) and the product extracted with DCM (5 x 10 mL). The solvent was evaporated under vacuum to yield acetylated sterols (30 mg). The derivative was analysed on a Diol HPLC column, using 5% DCM/hexane as eluent. The peaks were still overlapping.

7.2.4 Studies on HIV-inhibitory Components from a *Styela* sp.

Collection The ascidian was collected at the break-water, Lyttelton Harbour, Christchurch, New Zealand, in 1993 and belongs to the family *Styelidae,* order *Pleurogonia*

Extraction A frozen sample of the ascidian (100 g) was soaked in MeOH (200 mL) overnight, ground in a blender, and filtered. The residue was successively extracted with MeOH (2 x 80 mL), 50% MeOH/DCM (80 mL), and DCM (80 mL). The five extracts were combined, concentrated to an aqueous solution (about 100 mL), and absolute ethanol (100 mL) added to precipitate the polysaccharides. After filtration (the filter pad was washed with 3 x 30 mL EtOH), the EtOH was removed under vacuum. The resultant aqueous portion was partitioned against EtOAc (3 x 80 mL) to yield the organic (290 mg) and aqueous extracts (3.69 g).

Chemical screening for anti-HIV activity Sephadex G-25 and C4 Wide Pore (WP) cartridges were chosen for the aqueous, while C18, Diol and PEI WP were selected for the organic extract. The experimental procedures were the same as those described in Sections 7.2.1 and 7.2.3. The HIV-inhibitory activity was observed in three fractions (F1, F2 and F4 with inhibition % 37, 106 and 18, respectively) from the C4 column. The third fraction (F11) from C18, the second fraction (F13) from Diol, and the third fraction (F17) from PEI WP columns showed only trace activity (inhibition %<10). No activity was found in any fractions from the G-25 column.

Attempt at isolation of the anti-HIV compounds from *Styela* sp. A portion (200 mg) from the aqueous extract was dissolved in water and chromatographed on C-18 (8 g). This column was eluted initially with H_2O, then with MeOH/H_2O mixture (5% to 50%). The anti-HIV activity was observed in the first three fractions (inhibition% 17, 1 and 15 for F1, F2 and F3, respectively).

7.2.5 Chemical Screening of an Unidentified Sponge (code No. 9401-01)

Extraction The frozen sponge (550 g) was crushed to small pieces, soaked in MeOH (600 mL) for 3 h, and blended. Further extraction was achieved successively with MeOH (500 mL), 50% MeOH/DCM (400 mL), DCM (2 x 400 mL). After removing the organic solvents, the extract was partitioned between water and EtOAc. The combined EtOAc extract was evaporated to dryness to afford the organic extract (340 mg). The aqueous portion was freeze-dried to yield aqueous extract (2.04 g).

Chemical screening The chemical screening was carried out in the same way as described in Sections 7.2.1 and 7.2.3 for the organic and aqueous extracts, respectively. The best results were found for F2 and F3 with inhibition % 11 and 14 respectively.

7.2.6 Chemical Screening of Organic Extract from *Callyspongia irregularis*

C18, Diol and LH-20 cartridges were used for screening the organic extract using the same method as described in Section 7.2.1. The anti-viral and cytotoxic activity were evaluated by evaluation of the viral inhibition, or cytotoxic zones (+- for least active, 4+ for most active).[123] The antiviral activity was observed in the second (4+) and third (2+)

fractions from C18 column, the first (+) and the second (4+) fractions from Diol, and the second fraction (4+) from LH-20 column.

7.2.7 Chemical Screening of an Unidentified Sponge (code No. 95CR4-18)

Extraction Frozen sponge (5 g) was crushed to small pieces, soaked in MeOH/DCM (66%, 50 mL), blended, and left at room temperature for 3 h. The material was centrifuged at 3000 rpm for 20 min. The supernatant was evaporated under vacuum.

Chemical screening The organic extract was dissolved in MeOH/DCM (3:1) (50 mg extract/mL). A sample of this solution (100 μL) was coated onto celite (200 mg). The chemical screening was carried out in the same way as described in Sections 7.2.1, using C-18, Diol, PEI WP, CBA WP and LH-20 cartridges. The cytotoxicity (IC_{50}, dil.) was found in the fractions 2 (969) and 3 (526) from C18 column, fraction 3 (233) from Diol, fraction 2 (3688) from CBA WP, fraction 2 (<97.5) from LH-20. The active fraction from LH-20 was loaded onto a C18 cartridge, and eluted with the same solvent as used in the above procedure. The second fraction showed a weak cytotoxicity (IC_{50}=1057). The other two fractions from the same column were not cytotoxic (IC_{50}>12500).

7.2.8 Chemical Screening of an Unidentified Sponge (code No. 95CR2-10)

The primary screening on the aqueous extract from this sponge indicated that it was strongly cytotoxic against the P388 cell line (IC_{50} <97.5). A sample of the frozen sponge (5g) was extracted with water (50 mL), and analysed in the same way as described in Section 7.2.7. The aqueous extract was further analysed on C4 WP, PEI WP, CBA WP and Sephadex

G-25 cartridges using the same methods described in Sections 7.2.1 and 7.2.3. Only trace cytotoxicity (IC_{50}=3909 dil.) was recognised in the first fraction from CBA column, while the standard still exhibited intense activity (IC_{50} <97.5).

7.3 Work Related to Chapter 3

7.3.1 Studies on Antiviral and Antitumour Components of *Callyspongia irregularis*

Collection The sponge was collected by SCUBA diving at Lighthouse Reef Point, Moeraki, South Island, New Zealand. The genus *Callyspongia* belongs to family *Callyspongiidae*, order *Haplosclerida.*

Extraction of *Callyspongia irregularis* Frozen sponge (380g) was thawed in MeOH (600 mL), blended and filtered through a sintered funnel. The residue was in turn extracted with MeOH (600 mL), 50% MeOH/DCM (200 mL), and finally H_2O (600 mL). The extracts were combined and the organic solvent was removed under vacuum. To the remained aqueous portion absolute alcohol was added to make a 50% ethanol solution. The resultant solution plus precipitate was filtered off through celite. After evaporation of the ethanol under vacuum, the aqueous portion was freeze-dried to yield an extrcat (18.3g).

Isolation and purification of mycalamide A The organic extract (16 g) was suspended in 50 % MeOH/DCM, and filtered. The residue was washed with the same solvent until the colour of the filtrate was a very light straw colour. The total organic solvent was removed under vacuum, the residue redissolved in a minimum valume of (4 mL) 50 % MeOH/DCM and coated onto celite (13 g). This material was loaded onto the head of a column (9 cm i.d.), containing C18 (30 g). This column had been equilibrated to water, and was then eluted with water (200 mL). Step gradient elution using MeOH/H_2O, MeOH, DCM/MeOH, and DCM (50 mL/fraction) afforded 39 fractions. Finally, the column was washed with MeOH and DCM alternatively, producing a 40th

fraction. The fractions were combined to six new fractions based on both bioassay results and the TLC behaviour.

Fraction I.3 (Scheme 3.3.1) was separated by high-speed countercurrent chromatography (HSCCC), using petroleum ether/EtOAc/MeOH/H_2O (6:11:6:4), upper phase as mobile phase. A total of 26 fractions were collected from each run. The active fractions were then combined, and repeatedly separated with HSCCC using same solvent system as above. The antiviral activity was concentrated in fraction II.2 (Scheme 3.3.1).

This active fraction (3.9 mg) was further purified on a Diol cartridge (500 mg, Bond Elut 20H), eluted with a gradient system using EtOAc/DCM, MeOH/EtOAc, and MeOH (5 mL/fraction). The eluates were re-combined to five fractions based on bioassay results. The fraction III.2 (Scheme 3.3.1) mainly contained a highly active antiviral component. The ^{1}H NMR spectrum showed signals characteristic of mycalamide, as well as impurities. However, no further purification was made due to the limited amount of the material (about 200 μg) available.

Negative FABMS (NOBA) spectrum exhibited $(M-H)^-$ ion at *m/z* 502 (very weak). The molecular formula $C_{23}H_{41}NO_{10}$ was deduced from a comparison of the molecular weight with the HRFABMS of an abundant daughter ion at *m/z* 472.

^{1}H NMR data of mycalamide ($CDCl_3$) δ 7.49 (d, *J*=8.7 Hz, NH), 5.87 (t, *J*=9.9 Hz, H10), 5.13 (d, *J*=6.9 Hz, 10-OCH*R*), 4.87 (d, *J*=6.9 Hz, 10-OCH*S*), 4.85 (m, 4=CHZ), 4.74 (m,4=CHE), 4.30 (s, H7), 4.23 (dd, *J*=6.8, 7.3 Hz, H12), 3.99 (dq, *J*=2.7, 6.6 Hz, H2), 3.85 (dd, *J*=6.8, 9.7 Hz), 3.74 (m, H17), 3.62 (d, *J*=5.5 Hz, H15), 3.57 (dd, *J*=3.5, 11.3 Hz, H18), 3.55 (s, 13-OCH_3), 3.45 (d, *J*=10.3, H13), 3.38 (dd, *J*=6.2, 11.3 Hz,

H18), 3.30 (s, 6-OCH_3), 2.36 (m, H5), 2.24 (m, H3), 1.19 (d, *J*=2.6 Hz, 2-CH_3), 0.99 (d, *J*=7.0 Hz, 3-CH_3), 0.97 (s, 14-CH_3*R*), and 0.87 (s, 14-CH_3*S*).

Isolation of other compounds from *C. irregularis* Fraction II.1 (Scheme 3.3.1) was processed on a RPHPLC (C18) column, using 57% MeOH/H_2O (containing 0.05% TFA). There was no antiviral activity found in any of the fractions. A similar result was obtained when fraction II.5 was also chromatographed on the same system using 60% MeOH/H_2O (containing 0.05% TFA).

The cytotoxic components in fraction I.4 (Scheme 3.3.1) was concentrated in fraction IX.2 by HSCCC. The weakly cytotoxic fraction was purified further using RPHPLC, using 80% MeOH/H_2O (containing 0.05% TFA). No activity was detected in any of the subsequent fractions.

Fraction V.1, which displayed mild cytotoxicity, was loaded onto a column (containing 15 g LH-20). The column was eluted with DCM, and yielded two active fractions (fractions VI.4 and VI.5). These fractions were chromatographed on a Diol (3.6 g) column, which was eluted with DCM, and 2-30% MeOH/DCM. The IC_{50} values recorded for fractions VIII.1, 2 and 3 in a P388 bioassay were >12500, 2455, and >12500.

7.3.2 Isolation of Cytotoxic Compounds from *Callyspongia* sp. 2

Extraction of *Callyspongia* sp. 2 Frozen sponge (1.6 Kg) was steeped in MeOH (1.7 L) at room temperature for 3 h, then blended. The material was filtered through celite. The residue was in turn extracted with MeOH (1.2 L), 50% MeOH/DCM (1.2 L), and DCM (2x1.2 L), each for 1 h. The MeOH extracts were combined, and partitioned against

EtOAc after removal of the MeOH using a cyclone evaporator. The EtOAc extract was combined with the organic extracts to give the crude extract (810 mg).

Chemical screening of the organic extract Organic extract (corresponding to about 5 mg solid material) was coated onto celite (300 mg). C18, Diol and PEI WP cartridges (3 mL, 3 mL and 6 mL respectively) were chosen for the chemical screening. The screening was carried out in the same way as described in Section 7.2. The cytotoxic activity was distributed as following: >25000 (50% MeOH/H_2O), 6250 (MeOH), 2265 dil. (50 MeOH/DCM) from C18; >12500 (DCM), 3144 (EtOAc), >6500 dil. (MeOH) from Diol; >12500 (0.025 M NH_4Ac), >12500 (0.5 M HOAc), >25000 (0.5 M sodium citrate).

Isolation of the cytotoxic components The organic extract (810 mg) was coated onto celite, and loaded onto a C18 column (8 g). The column was eluted with a gradient mode using H_2O, MeOH/H_2O, MeOH, and DCM. A total of 16 fractions were collected. They were combined to just 6 fractions based on bioactivity. Fraction I.4 (Scheme 3.4.1), the most active fraction (IC_{50} 31 ng/mL), was processed further by RPHPLC, using 85% MeOH/H_2O as eluent. The resultant bioactive fractions were combined to give fraction V.3 (Scheme 3.4.1), which was then combined with fraction II.3 (Scheme 3.4.1) from the previous Diol column. The material was dissolved in DCM, and chromatographed on an LH-20 column (2 x 60 cm, 30 g) equilibrated to DCM. Eleven fractions were collected from the column, which was eluted with DCM (F1-9), 2% MeOH/DCM (F10), then washed with MeOH and DCM (F11). The active components were contained in fractions 8-10, which were combined to afford fraction VI.3 (Scheme 3.4.1). The less active material (F3-7) formed VI.2. The ^1H NMR spectrum of the fraction VI.3 indicated that it

was still a complex mixture. To purify the active compounds, normal phase Diol HPLC was used. The column was eluted with a gradient using IPA/hexane from 5% to 40%. Good resolution was achieved (a total of 10 peaks were resolved and collected individually). However, the most potent IC_{50} was just 1120 ng/mL, implied a major loss of the activity. Fractions having similar activity were combined.

Fraction I.3 (Scheme 3.4.1) was loaded onto a Diol column (14g), which was eluted with a gradient of DCM/petroleum ether from 70% to 100%, and MeOH/DCM from 2% to 10%. The active fraction (II.3) showed similar ^1H NMR spectrum and IC_{50} value to V.3 (Scheme 3.4.1). These two fractions were combined. A less active fraction (II.4) was also chromatographed on Diol, eluting with 50% MeOH/DCM. There was no increase in the bioactivity observed, while a decrease in bioactivity was recognised in the bioassay results of the eluates from a RPHPLC column.

Fraction I.5 (Scheme 3.4.1) was processed by loading it onto a Diol column (2.5x60 cm, 18 g). A solvent gradient from 40% petroleum ether/DCM to 10% MeOH/DCM was used in the chromatography. The cytotoxicity was concentrated in fractions 11-14, which were combined to yield F VIII.3. Before a Lobar column was chosen for further purification, test runs were carried out on the fraction using the RPHPLC (ODS column) solvent systems. The bioactivity was found to spread in a wide range of fractions in all cases. So, all collections were combined, and chromatographed instead on a Diol Lobar column with IPA/hexane (30%) as eluent. The active components were still distributed, but the first and last fractions did not contain any activity. This suggested the active fraction must be a complex mixture. After passage through a Diol cartridge, the active fraction (IX.2, in Scheme 3.4.1) was divided into

three parts. The ^{1}H NMR spectrum of the most active one (X.1) showed that it contained at least one major as well as several minor components (see Fig. 3.4.1).

7.4 Work Related to Chapters 4-6

7.4.1 Common Methods

Collection *Lamellomorpha strongylata* was collected in February 1995, and recollected in April 1996 by benthic dredging (-80 to -100 m) from the Mernoo Bank, Chatham Rise, New Zealand. A voucher specimen is held in the museum at NIWA, Wellington, New Zealand.

Extraction and Isolation The extraction of the sponge and separation of the theonellapeptolide fraction from the calyculin and swinholide fractions has been reported previously.[118] Final purification of the peptolides was carried out by reverse-phase HPLC on a Philips PU4100 using an ODS column (4.6 mm x 25 cm) eluting with H_2O (0.1% TFA)/MeOH (0.1% TFA) (11:89). A limited sample (2 mg) of the crude peptide was injected in each run. Cutting the appropriate peak in multiple runs gave pure theonellapeptolide IIIe (**50**) (94 mg).

Large Scale Separation of Theonellapeptolides: To obtain enough quantity of the minor peptides for structural studies, a larger scale separation was carried out. Frozen sponge (5.96 kg) was ground and extracted with MeOH/DCM (3:1, 2 x 6L). The combined extracts were concentrated to an aqueous suspension, and partitioned against dichloromethane (3 x 750 mL). After removal of solvent, the organic extract were coated onto ODS silica, loaded onto a C18 column (90 g) and eluted using a steep, step-gradient pattern from 100% water to 100% methanol, followed by stripping the column with 50% DCM/MeOH and 100% DCM. The cytotoxic fractions were combined and chromatographed on a Diol column with a step gradient elution pattern using hexane-dichloromethane mixtures. The fractions eluting at

dichloromethane (70-100%) were found to contain theonellapeptolides. The crude peptolides (2.6g) were dissolved in 50% MeOH/DCM (5 mL), and chromatographed on LH-20 (3m length, 200 cm id., 250g). Elution of the column with the same solvent afforded 23 fractions. Combination of fractions 7-17, based on both the bioactivity and Rf value on TLC, yielded a mildly cytotoxic fraction (1.83 g). The ^{1}H NMR showed that the majority of the mixture was the peptolides. This fraction was further chromatographed on silica gel (40 g) equilibrated to 5% hexane/EtOAc. Elution with 5% hexane/EtOAc, followed by EtOAc and 90% EtOAc/MeOH (20 mL/fraction), produced 22 fractions. Fractions 16-22 were combined according to their Rf values on TLC and their ^{1}H NMR spectra, which yielded the peptolide mixture (1.76 g). The peptolides were isolated and purified by semipreparative C18 HPLC column (10 mm x 2 5 cm, 5 μm, 100Å) using a Shimadzu LC-4A solvent delivery system. The column was eluted with 10% H_2O/MeOH (containing 0.1% TFA) in an isocratic mode. A total of ten fractions were collected. Of these, eight contained peptolide components (including the mass from the analytical column) of IIIa (24 mg), IIIb (20 mg), IIIc (20 mg), IIId (113 mg), IIIe (1.3g), IIIg (84 mg), IIIh (26 mg), IIIf (2 mg).

Acid Hydrolysis of Peptolides A sample of each theonellapeptolide (1 mg) was placed in an hydrolysis tube and HCl (6 M) (1 mL) was added. After sealing under N_2, the tube was heated at 110 °C for 17 hrs. In each case, the acid solution was evaporated to dryness under vacuum to give the product.

Amino Acid Analysis by GC/MS The *N*-trifluoroacetamide-*n*-butyl esters of the amino acids were prepared under standard conditions.[145] The acid hydrolysate of each theonellapeptolide (1 mg)

was transferred to a Reacti-vial (1 mL) and dried under vacuum before HCl-*n*-BuOH (2 M) (400 μL) was added. The vial was sealed with a Teflon-lined cap and, after sonication (15 s), was heated at 110°C in an oil bath (10 min), followed by further sonication (10 s) and heating (20 min). After cooling the solvent was evaporated under vacuum. A trifluoroacetic anhydride-DCM mixture (750 μL) (1:2) was added to each sample, which was then heated (150°C) for 5 min. The remaining reagent and solvent were removed under N_2 at 50°C and the product then dissolved in dry DCM (600 μL) for GC analysis. The GC-MS analysis of each sample was performed on a DB Wax column (30 m x 0.25 mm) with an initial column temperature of 50°C which was held for 15 min post-injection, then increased to 130°C at 1°/min, to 250°C at 23°/min and finally held at 250°C for 20 min. The individual amino acids were identified on the basis of retention time against standards, and by analysis of the MS data.

Methanolysis of the theonellapeptolides Each theonellapeptolide (~1 mg) was dissolved in dry MeOH (220 μL) with NaOMe/MeOH (4%) (66 μL). The MeOH solution of each theonellapeptolide was stirred at room temperature for 3 h before pouring into ice water and extraction with EtOAc. Evaporation of the solvent gave the products (~0.9 mg) which were purified by reverse-phase HPLC [H_2O (0.1% TFA)/MeOH; 18:82].

7.4.2 Synthesis of Amino Acid Derivatives

Synthesis of N-methyl L-alanine (60) L-Ala (268 mg) was dissolved in 2M NaOH (1.5 mL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (1.04 mL) and 2M NaOH (1.7 mL) were added alternatively, while the mixture was vigorously stirred in the ice

bath. After continued stirring at room temperature for another 2 h, the solution was adjusted to pH 10 with NaOH. The basic solution was partitioned against ether, then acidified to pH 2 with 6M HCl, before extraction with the ether (4 x 4 mL). The solvent was removed under vacuum to yield benzyloxycarbonyl-L-Ala (584 mg).

Benzyloxycarbonyl-L-Ala (3 mmol) was dissolved in tetrahydrofuran (THF) (2 mL) in a reacti-vial. To this solution CH_3I (200 µl) was added. After the solution was cooled to 0°C in an ice-bath NaH (40 mg) was added to the vial. The reaction mixture was stirred under N_2 at room temperature for 24 h. The remained NaH was destroyed by adding water, and the resulting solution was partitioned against ether (3 x 3 mL). The aqueous portion was acidified to pH 2 with 6N HCl, and extracted with EtOAc (4 x 3 mL). The ethyl acetate extract was washed with 5% sodium thiosulfate (2 x 1 mL) and water (2 x 3 mL), and dried over $MgSO_4$. The solvent was removed under vacuum. The product (591 mg) was purified on a Microsorb-MV C18 column using 50% MeOH/H_2O as eluent.

The benzyloxycarbonyl N-methyl alanine (587 mg) was vigorously stirred with 10% Pd-C (20 mg) in ethanol (2 mL) under hydrogen at room temperature until the CO2 evolution ceased. After the residual hydrogen was replaced by N_2, the catalyst was filtered off over celite under N_2. The solvent was removed to give a crude product (158 mg). The residue was dissolved in water, and passed through a C18 cartridge, which was eluted with H_2O. The aqueous eluate was dried under high vacuum to yield pure L-MeAla (153 mg).

^{1}H NMR (D_2O) δ 3.43 (1H, q, *J*=7.2 Hz, H2), 2.51 (3H, s, NCH_3), 1.29 (2H, t, *J*=7.2 Hz, H3).

Synthesis of N-methyl-β-alanine (61) βAla (198 mg) was dissolved in 2M NaOH (1.0 mL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (930 μL) and 2M NaOH (1.5 mL) were added alternatively, while the mixture was vigorously stirred in the ice bath. N-methylation was carried out as for L-MeAla (see above) to yield N-methyl-β-alanine (124 mg).

^{1}H NMR (D_2O) δ 2.92 (2H, t, J=6.6, 4.4 Hz, H3), 2.45 (3H, s, NCH_3), 2.29 (2H, t, J=6.9, 6.3 Hz, H2).

Synthesis of N-methyl-L-isoleucine (62) L-Ile (262 mg) was dissolved in 2M NaOH (1.5 mL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (700 μL) and 2M NaOH (1.2 mL) were added alternatively, while the mixture was vigorously stirred in the ice bath. N-methylation was carried out as for L-MeAla (see above) to yield N-methyl-L-isoleucine (144 mg).

^{1}H NMR (10% CD_3OD/D_2O) δ 3.33 (1H, d, *J*=3.9 Hz, H2), 2.55 (3H, s, NCH_3), 1.78 (1H, m, H3), 1.39 (1H, m, H4), 1.15 (1H, m, H4), 0.81 (3H, d,d, *J*=7.2, 6.9Hz, H5), 0.83 (3H, d, *J*=7.2 Hz, H6).

Synthesis of N-methyl-D-isoleucine (63) D-Ile (29 mg) was dissolved in 2M NaOH (400 μL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (60 μL) and 2M NaOH (600 μL) were added alternatively, while the mixture was vigorously stirred in the ice bath. N-methylation was carried out as for L-MeAla (see above) to yield N-methyl-D-isoleucine (17 mg).

^{1}H NMR (10% CD_3OD/D_2O) δ 3.35 (1H, d, *J*=3.9 Hz, H2), 2.57 (3H, s, NCH_3), 1.80 (1H, m, H3), 1.40 (1H, m, H4), 1.16 (1H, m, H4), 0.82 (3H, d,d, *J*=7.5, 6.9 Hz, H5), 0.84 (3H, d, *J*=7.5 Hz, H6).

Synthesis of N-methyl-L-*allo*isoleucine (64) L-*allo*Ile (14 mg) was dissolved in 2N NaOH (200 μL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (30 μL) and 2N NaOH (400 μL) were added alternatively, while the mixture was vigorously stirred in the ice bath. N-methylation was carried out as for L-MeAla (see above) to yield N-methyl-L-*allo*isoleucine (8 mg).

^{1}H NMR (D_2O) δ 3.38 (1H, d, *J*=3.9 Hz, H2), 2.59 (3H, s, NCH_3), 1.87 (1H, m, H3), 1.36 (1H, m, H4), 1.17 (1H, m, H4), 0.84 (3H, d,d, *J*=7.5, 7.2 Hz, H5), 0.86 (3H, d, *J*=7.2 Hz, H6).

Synthesis of N-methyl-D-*allo*isoleucine (65) D-*allo*Ile (100 mg) was dissolved in 2M NaOH (500 μL) and stirred in an ice-water bath for 20 min. Benzyl chlorocarbonate (300 μL) and 2M NaOH (500 μL) were added alternatively, while the mixture was vigorously stirred in the ice bath. N-methylation was carried out as for L-MeAla (see above) to yield N-methyl-D-*allo*isoleucine (57 mg).

^{1}H NMR (D_2O) δ 3.49 (1H, d, *J*=3.9 Hz, H2), 2.72 (3H, s, NCH_3), 2.00 (1H, m, H3), 1.49 (1H, m, H4), 1.30 (1H, m, H4), 0.97 (3H, d,d, *J*=10.2, 7.5 Hz, H5), 0.99 (3H, d, *J*=7.5 Hz, H6).

7.4.3 Theonellapeptolide IIIe (50)

Theonellapeptolide IIIe (50) colourless plates (acetone/H_2O: 1:1): MP 184-186°; $[\alpha]^{22}_D$ -48.6° (c 1.0, MeOH); IR ($CHCl_3$) ν_{max} 3352, 3006, 2966, 2935, 1734, 1672, 1628, 1533, 1468, 1173 cm^{-1}; ^{1}H NMR (CD_3OD/C_6D_6, 2:8) see Table 4.1; ^{13}C NMR (CD_3OD/C_6D_6, 2:8) see Table 4.1; HRFABMS, *m/z* 1550.8577 $[(M+Cs)^+$, -0.1 mmu for $C_{71}H_{127}N_{13}O_{16}Cs$). FABMS, *m/z* 1419 $(M+H)^+$, 1248, 1219, 1206,

1121, 1106, 1038, 924, 810, 712, 694, 655, 525, 496, 439, 380, 299, 269, 211, 182, 144, 100, 86, 72, 58, 44.

Methyl Ester of Theonellapeptolide IIIe (51) HRFABMS, *m/z* 1582.8860 [(M+Cs)$^+$, +2.0 mmu for $C_{72}H_{131}N_{13}O_{17}Cs$). FABMS, *m/z* 1451 (M+H)$^+$, 1421, 1280, 1152, 1107, 1051, 1022, 966, 926, 895, 867, 740, 711, 669, 654, 615, 584, 556, 527, 496, 485, 429, 400, 369, 344, 299, 277, 185, 172, 100, 72, 57, 44.

Methanolysis of 50 Theonellapeptolide IIIe (**50**) (9 mg) was dissolved in dry MeOH (2 mL) with NaOMe/MeOH (4%) (0.6 mL). After incubation for 3 h, the sample was treated in the same way as described in Section 7.4.1 to give the methanolysis product **51** (6 mg).

X-ray Structure Determination of Theonellapeptolide IIIe

A colourless crystal of theonellapeptolide IIIe **50,** obtained as a polyhydrate from acetone-water (1:1), was used in the data collection. Data reported here were derived from an irregular chunk of dimensions (0.6 x 0.4 x 0.15 mm^3). Crystal data: formula $C_{71}H_{127}N_{13}O_{16}/(H_2O)_{18}$, FW=1418/(324) ; orthorhombic, $P2_12_12_1$, a=12.507(3), b=19.573(4), c=41.416(8) Å, V=10,139 Å^3, Z=4, Rho=1.142.

The data collection nominally covered a hemisphere. Intensities of 18,590 unique reflections were collected on a Siemens SMART area detector system, fitted with a nitrogen low-temperature gas flow device, using MoKα (λ=0.71071 Å) X-radiation. These data were processed using program SAINT which corrects for LP effects and crystal decay for the duration of the experiment.

The structure was solved by direct methods and refined using the SHELXTL system of programs and all the unique intensity data. The best converged conventional *R*-factor was 0.12 for 12,029 refractions with $F_0>4\sigma$ (F_0). All numerical data discussed here came from programs in the SHELXTL system as well as the perspective diagram of the molecule (Figure 4.6.1), which defines the atom labelling scheme used for the crystal structure.

Determination of Absolute Configuration A sample of the acid hydrolysate of theonellapeptolide IIIe **50** (1 mg) was dissolved in aq. Li_2CO_3 (0.04M) (3 mL) and reacted with dansyl chloride (Dan-Cl 2.2 mg/AcCN 1mL) (2.3 mL)[158] at room temperature (2 h). The solvent was removed under vacuum. The Dan-amino acids were further purified on a Diol cartridge with DCM-ether-HOAc (5:1:1) as eluent. The first fraction (0.8 mL) was collected and dried under N_2 (0.8 mg). The Dan-derivatives were analysed on a reverse-phase HPLC column (ODS, 4.6 x 220 mm) run isocratically using L-Val (4 mM)-copper acetate (2 mM) -sodium acetate (0.3 M) (pH 7) as chiral eluent (1 mL/min). Standard Dan-L-Thr and Dan-DL-Thr were coinjected with the dansyl derivatives. The retention times of the standard Dan-D-Thr and Dan-L-Thr were 720 and 780 s, respectively. The Dan-Thr in the hydrolysate eluted at 780 s, which indicated an L-configuration for this residue.

7.4.4 Theonellapeptolide IIIb (52)

^1H and ^{13}C NMR data in Table 5.1; a colourless powder; $[\alpha]_D$ -49° (c 0.7, MeOH); IR ($CHCl_3$) ν_{max} 3352, 3024, 3008, 2966, 2936, 2878, 2484, 1734, 1672, 1626, 1522, 1464, 1416, 1117, 1099, 1045 cm^{-1}; HRFABMS *m/z* 1508.8134 ($[M+Cs]^+$, +2.5 mmu for $C_{68}H_{121}N_{13}O_{16}Cs$); FABMS (glycerol) *m/z* 1377 $(M+H)^+$, 1304, 1205, 1120, 1106, 1037, 923, 767,

708, 669, 651, 623, 524, 496, 468, 380, 368, 340, 312, 257, 211, 144, 100, 86, 72, 58, 44.

Amino Acid Analysis by GC/MS The *N*-trifluoroacetamide-*n*-butyl esters of the hydrolysate were prepared and analysed as described in Section 7.4.1.

Methanolysis of 52 Theonellapeptolide IIIb (**52**) (1 mg) was dissolved in dry MeOH (220 μL) with NaOMe/MeOH (4%) (66 μL). For details see Section 7.4.1. The methanolysis product **53** (0.9 mg) was produced.

Methyl Ester of Theonellapeptolide IIIb (53) FABMS, *m / z* 1447 $(M+K)^+$, 1431 $(M+Na)^+$, 1409 $(M+H)^+$, 1237, 1152, 1065, 1051, 1037, 1011, 980, 966, 867, 853, 740, 669, 654, 556, 527, 496, 429. 400, 344, 337, 299, 223, 172, 100, 72, 58, 44.

Determination of Absolute Configuration:

GITC method A solution of standard amino acids, or a theonellapeptolide hydrolysate (20 μL; ~5.05 mg/mL each) was combined with triethylamine (6%, 10 μL) and 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl isothiocyanate (GITC; Aldrich 33.858-3) (1%) in acetone (20 μL) and stood at room temperature for 5 min. An aliquot of the product (10 μL) was analysed on an LC/MS system using an ODS column (150 x 4.6 mm). The identity and chirality of the individual amino acids were identified by comparison of their mass and retention time prepared from GITC derivatives against those of the optically pure standard samples.

FDAA method A standard mixture of amino acids was prepared (0.05 mg/mL for each amino acid in 0.05% TFA). An aliquot of this amino acid mixture (20 μL), or the theonellapeptolide hydrolysate (20 μL) was combined with triethylamine (6%, 10 μL) and Nα-(2,4-Dinitro-5-fluorophenyl)-L-alaninamide (FDAA; Aldrich 36,605-6) in acetone (1%). This mixture was incubated at 40°C for 1 hour in a microvial (100 μL), then diluted with water (50 μL). The product was analysed as above. (GITC and FDAA derivatives were prepared and analysed by Dr, Shegi and Dr. Pannell at NIDDK/NIH, USA).

7.4.5 Structural Study on Theonellapeptolide IIIa (54)

Theonellapeptolide IIIa (54): a colourless powder; $[\alpha]_D$ -40° (c 0.8, MeOH); IR ($CHCl_3$) ν_{max} 3346, 3030, 3014, 2935, 1734, 1670, 1643, 1624, 1522, 1466, 1416, 1306, 1196, 1140, 1099, 1049 cm^{-1}; HRFABMS *m/z* 1584.8110 ($[M+Cs]^+$, mmu for $C_{70}H_{125}N_{13}O_{17}SCs$); FABMS (glycerol) *m/z* 1453 $(M+H)^+$, 1281, 1252, 1187, 1120, 923, 843, 776, 745, 654, 612, 569, 524, 495, 470, 439, 397, 380, 333, 312, 269, 224, 196, 172, 144, 100, 86, 72, 58, 44; NMR data: δ_H (35% $CD_3OD/CDCl_3$) methoxyl acetate [3.69 (2H, s, H2), 3.21 (3H, s, OCH_3)], Val 1 [7.07 (1H, d, *J*=8.7 Hz, NH), 4.62 (1H, m, H2), 1.84 (1H, m, H3), 0.73 (3H, CH_3), 0.68 (3H, CH_3)], MeHyMet [4.54 (1H, H2), 2.80 and 2.39 (1H, 1H, H4), 2.50 (3H, s, NCH_3), 2.38 (3H, s, SCH_3)], Thr [8.11 (1H, brs, NH), 4.94 (1H, m, H3), 4.58 (1H, m, H2), 0.79 (1H, d, *J*=6 Hz, H4)], Me-β-Ala [3.74 (1H, m, H3), 3.02 (1H, m, H3), 2.69 (3H, s, NCH_3), 2.23 (1H, m, H2), 2.10 (1H, m, H2)], Val 2 [8.12 (1H, brs, NH), 4.99 (1H, m, H2), 1.41 (1H, m, H3)], MeIle 1 [4.80 (1H, d, *J*=11.1 Hz, H2), 3.08 (3H, s, NCH_3), 1.98 (1H, m, H3), 1.17 (1H, m, H4), 0.84 (3H, H5), 0.64 (3H, d, *J*=4.2 Hz, H6)], β-Ala [6.98 (1H, brs, NH), 3.68 (1H, m, H3), 2.97 (1H,

m, H3), 2.24 (1H, m, H2), 2.00 (1H, m, H2)], Ile [7.86 (1H, d, *J*=9.0 Hz, NH), 4.62 (1H, m, H2), 1.36 (1H, m, H3), 1.13 (1H, m, H4), 0.71 (3H, H5), 0.69 (3H, H6)], MeIle 2 [4.58 (1H, m, H2), 3.08 (3H, s, NCH_3), 2.38 (1H, m, H3), 1.12 (1H, m, H4), 0.71 (3H, H5), 0.52 (3H, d, *J*=6.3 Hz, H6)], MeAla [4.72 (1H, m, H2), 2.99 (3H, s, NCH_3), 1.18 (3H, d, *J*=6.6 Hz, H3)], Ala [6.77 (1H, d, *J*=4.5 Hz, NH), 4.04 (1H, m, H2), 0.94 (3H, d, *J*=5.7 Hz, H3)], Leu [7.52 (1H, t, *J*=6.0 Hz, NH), 4.63 (1H, m, H2), 1.71 (1H, m, H3), 1.60 (1H, m, H3)], MeLeu [4.32 (1H, m, H2), 2.94 (3H, s, NCH_3)].

δ_C (35% $CD_3OD/CDCl_3$) methoxyl acetate [169.6 (s, C1), 71.0 (t, C2), 58.6 (s, OCH_3)], Val 1 [171.5 (s, CO), 53.7 (d, C2), 31.7 (d, C3)], MeHyMet [169.7 (s, C1), 60.7 (d, C2), 56.2 (d, C3) 50.6 (t, C4), 37.4 (s, SCH_3)], Thr [168.2 (s, CO), 70.0 (d, C3), 52.3 (d, C2), 16.7 (q, C4)], MeβAla [171.0 (s, CO), 44.3 (t, C3), 35.0 (q, NCH_3), 33.5 (t, C2)], Val 2 [174.9 (s, CO), 52.3 (d, C2), 37.6 (d, C3)], MeIle 1 [171.7 (s, CO), 55.1 (d, C2), 32.9 (d, d3), 30.9 (q, NCH_3), 15.0 (q, C6)], β-Ala [173.3 (s, CO), 34.7 (t, C3), 33.5 (t, C2)], Ile [174.9 (s, CO), 54.1 (d, C2), 39.5 (d, C3), 11.2 (q, C5) 20.3 (q, C6)], MeIle 2 [173.4 (s, CO), 52.3 (d, C2), 37.9 (q, NCH_3), 37.6 (d, C3), 25.8 (t, C4), 14.1 (q, C6)], MeAla [169.7 (s, CO), 56.2 (d, C2), 34.7 (q, NCH_3), 13.9 (q, C3)], Ala [173.1 (s, CO), 50.6 (d, C2), 17.3 (q, C3)], Leu [172.6 (s, CO), 47.7 (d, C2), 36.6 (t, C3)], MeLeu [173.7 (s, CO), 60.7 (d, C2), 30.6 (q, NCH_3)].

Amino Acid Analysis by GC-MS The *N*-trifluoroacetamide-*n*-butyl esters of the hydrolysate were prepared and analysed as described in Section 7.4.1.

Methanolysis of 54 Theonellapeptolide IIIa (**54**) (1 mg) was dissolved in dry MeOH (220 μL) with NaOMe/MeOH (4%) (66 μL). The

methanolysis product **55** (1.1 mg) was isolated and purified as described in Section 7.4.1..

Methyl Ester of Theonellapeptolide IIIa (55) FABMS, *m / z* 1485 $(M+H)^+$, 1141, 1056, 1051, 867, 787, 745, 740, 695, 679, 603, 581, 521, 489, 429. 400, 375, 337, 315, 223, 172, 131, 100, 72, 57, 45.

Determination of Absolute Configuration The configurations of the component amino acids of theonellapeptolide IIIa were determined in the same way as described in Section 7.4.4.

7.4.6 Structural Study of Theonellapeptolide IIIc (56)

Theonellapeptolide IIIc (56): a colourless powder; $[\alpha]_D$ -41° (c 0.8, MeOH); IR ($CHCl_3$) ν_{max} 3346, 3015, 2966, 2936, 2878, 2361, 1734, 1672, 1627, 1522, 1466, 1416, 1138, 1099 cm^{-1}; HRFABMS *m/z* 1522.8339 ($[M+Cs]^+$, +7 mmu for $C_{69}H_{123}N_{13}O_{16}Cs$); FABMS (glycerol) *m/z* 1391 $(M+H)^+$, 1318, 1219, 1120, 1106, 1092, 1037, 923, 795, 781, 726, 654, 640, 581, 524, 496, 468, 439, 425, 397, 380, 368, 312, 299, 285, 271, 211, 199, 182, 172, 144, 100, 86, 72, 58, 44; NMR data: δ_H (CD_3OH) methoxyl acetate [3.81 (2H, s, H2), 3.30 (3H, s, OCH_3)], Val [7.41 (1H, d, *J*=8.7 Hz, NH), 4.70 (1H, m, H2), 1.94 (1H, m, H3), 0.80 (3H, H4), 0.76 (3H, H5)], Leu 1 [7.99 (1H, d, *J*=4.5 Hz, NH), 4.74 (1H, m, H2), 1.79 (1H, m, H3), 1.15 (1H, m, H4), 0.79 (3H, H5)], Thr [8.86 (1H, d, *J*=6 Hz, NH), 5.10 (1H, m, H3), 4.59 (1H, m, H2), 0.92 (1H, d, *J*=4.2 Hz, H4)], MeβAla [4.42 (1H, m, H3), 2.68 (1H, m, H3), 2.65 (3H, s, NCH_3), 2.46 (1H, m, H2), 2.08 (1H, m, H2)], MeAla 1 [4.93 (1H, m, H2), 3.14 (3H, s, NCH_3), 1.29 (3H, d, *J*=4.2 Hz, H3)], MeIle 1 [4.91 (1H, d, *J*=7.8 Hz, H2), 3.09 (3H, s, NCH_3), 2.03 (1H, m, H3), 1.30 (1H, m, H4), 0.95 (3H, H5), 0.75 (3H, d, *J*=7.5 Hz, H6)], β-Ala [7.52 (1H,

dd, *J*=9.0, 5.4 Hz, NH), 4.09 (1H, m, H3), 2.98 (1H, m, H3), 2.23 (1H, m, H2), 2.03 (1H, m, H2)], Ile [8.48 (1H, d, *J*=6.0 Hz, NH), 5.05 (1H, m, H2), 1.55 (1H, m, H3), 1.00 (2H, m, H4), 0.87 (3H, H5), 0.66 (3H, H6)], MeIle 2 [4.99 (1H, m, H2), 3.17 (3H, s, NCH_3), 2.08 (1H, m, H3), 1.24 (2H, m, H4), 0.73 (3H, H5), 0.74 (3H, d, *J*=7.5 Hz, H6)], MeAla 2 [4.92 (1H, m, H2), 2.60 (3H, s, NCH_3), 1.23 (3H, d, *J*=4.5 Hz, H3)], Ala [7.23 (1H, d, *J*=4.2 Hz, NH), 4.03 (1H, m, H2), 1.09 (3H, d, *J*=4.5 Hz, H3)], Leu 2 [8.13 (1H, t, *J*=4.8 Hz, NH), 4.72 (1H, m, H2), 1.58 (1H, m, H3), 1.16 (1H, m, H3)], MeLeu [3.41 (1H, m, H2), 3.20 (3H, s, NCH_3), 1.70 (2H, m, H3), 1.49 (1H, m, H4), 0.93 (3H, H5), 0.84 (3H, H6)].

δ_c (35% $CD_3OD/CDCl_3$) methoxyl acetate [171.7 (s, C1), 72.4 (t, C2), 59.6 (s, OCH_3)], Val [174.7 (s, CO), 55.8 (d, C2), 32.1 (d, C3), 19.4 (q, C4), 17.8 (q, C5)], Leu 1 [171.7 (s, CO), 48.7 (d, C2)], Thr [169.9 (s, CO), 71.2 (d, C3), 54.0 (d, C2)], MeβAla [175.6 (s, CO), 45.7 (t, C3), 35.2 (q, NCH_3), 34.7 (t, C2)], MeAla 1 [174.0 (s, CO), 56.4 (d, C2), 32.1 (q, NCH3), 15.1(q, C3)], MeIle 1 [61.0 (d, C2), 32.4 (d, C3), 31.6 (q, NCH_3), 15.0 (q, C6), 9.9 (q, C5)], β-Ala [173.2 (s, CO), 35.8 (t, C2), 35.4 (t, C3)], Ile [176.4 (s, CO), 54.2 (d, C2), 39.1 (d, C3), 27.3 (t, C4), 11.7 (q, C5), 15.7 (q, C6)], MeIle 2 [173.7 (s, CO), 56.4(d, C2), 34.6 (d, C3), 31.9 (q, NCH_3), 15.5 (q, C6)], MeAla 2 [172.2 (s, CO), 58.2 (d, C2), 30.0 (q, NCH_3)], Ala [175.6 (s, CO), 52.9 (d, C2), 18.3 (q, C3)], Leu 2 [175.1 (s, CO), 49.2 (d, C2), 40.8 (t, C3)], MeLeu [172.7 (s, CO), 63.6 (d, C2), 38.9 (q, NCH_3), 38.8 (t, C3), 25.7 (d, C4), 23.5 (q, C5), 22.4 (q, C6)].

Amino Acid Analysis by GC/MS The *N*-trifluoroacetamide-*n*-butyl esters of the hydrolysate were prepared and analysed as described in Section 7.4.1.

Methanolysis of 56 Theonellapeptolide IIIc (**56**) (1 mg) was dissolved in dry MeOH (220 μL) with NaOMe/MeOH (4%) (66 μL). The treatment was the same as in Section 7.4.1. Methanolysis product **57** (1.0 mg) was produced.

Methyl Ester of Theonellapeptolide IIIc (57) FABMS, *m / z* 1423 $(M+H)^+$, 1377, 1251, 1079, 1037, 994, 867, 669, 683, 654, 563, 521, 471, 429, 343, 337, 315, 223, 100, 72, 57, 45.

Determination of Absolute Configuration The configurations of the component amino acids of theonellapeptolide IIIc were determined in the same way as described in Section 7.4.4.

7.4.7 Structural Study of Theonellapeptolide IIId (58)

Theonellapeptolide IIId (58): a colourless powder; $[\alpha]_D$ -47° (c 1.4, MeOH); IR ($CHCl_3$) ν_{max} 3352, 3026, 3007, 2966, 2936, 2876, 1734, 1674, 1626, 1526, 1468, 1416, 1467, 1416, 1117, 1092, 1053 cm^{-1}; HRFABMS *m/z* 1536.8409 ($[M+Cs]^+$, -1.2 mmu for $C_{70}H_{125}N_{13}O_{16}Cs$); FABMS (glycerol) *m/z* 1405 $(M+H)^+$, 1333, 1204, 1037, 1023, 909, 809, 795, 711, 654, 640, 581, 524, 496, 468, 439, 425, 397, 380, 368, 312, 299, 285, 211, 199, 182, 172, 144, 100, 86, 72, 58, 44; NMR data: δ_H (CD_3OH) methoxyl acetate [3.85 (2H, s, H2), 3.19 (3H, s, OCH_3)], Val 1 [7.48 (1H, d, *J*=8.7 Hz, NH), 4.98 (1H, m, H2), 1.99 (1H, m, H3), 0.90 (3H, H4)], MeLeu 1 [5.26 (1H, m, H2), 3.19 (3H, s, NCH_3), 2.03 (1H, m, H3), 1.45 (1H, m, H3), 1.45 (1H, m, H4), 1.03 (3H, H5), 0.79 (3H, H6)], Thr [9.03 (1H, d, *J*=6 Hz, NH), 5.42 (1H, m, H3), 4.94 (1H, m, H2), 1.27 (1H, br, H4)], Meβ Ala [4.68 (1H, m, H3), 2.72 (1H, m, H3), 2.72 (3H, s, NCH_3), 2.53 (1H, m, H2), 2.29 (1H, m, H2)], Val 2 [8.18 (1H, br, NH), 5.02 (1H, m, H2), 2.01 (1H, m, H3), 0.96 (3H, H4)], MeVal [5.22 (1H,

H2), 3.24 (3H, s, NCH_3), 2.05 (1H, m, H3), 1.13 (1H, H4), 1.06 (3H, H5)], β-Ala [7.62 (1H, d, brs, NH), 4.54 (1H, m, H3), 3.43 (1H, m, H3), 2.68 (1H, m, H2), 2.48 (1H, m, H2)], Ile [8.74 (1H, br, NH), 5.33 (1H, m, H2), 1.72 (1H, m, H3), 1.41 (1H, m, H4), 1.18 (1H, m, H4), 1.03 (3H, H5), 0.96 (3H, H6)], MeIle [5.18 (1H, m, H2), 3.31 (3H, s, NCH_3), 2.23 (1H, m, H3), 1.36 (2H, m, H4), 0.81 (3H, d, *J*=4.8 Hz, H6)], MeAla 2 [5.18 (1H, m, H2), 2.68 (3H, s, NCH_3), 1.42 (3H, d, *J*=6.0 Hz, H3)], Ala [7.41 (1H, brs, NH), 4.48 (1H, m, H2), 1.27 (3H, br, H3)], Leu [8.38 (1H, d, *J*=7.5 Hz, NH), 5.07 (1H, m, H2), 1.85 (1H, m, H3), 1.29 (1H, m, H3), 1.85 (1H, m, H4), 1.06 (3H, H5), 0.94 (3H, H6)], MeLeu 2 [3.69 (1H, m, H2), 3.26 (3H, s, NCH_3), 1.97 (2H, m, H3), 1.63 (1H, m, H4), 1.04 (3H, H5), 0.98 (3H, H6)].

δ_C (35% $CD_3OD/CDCl_3$) methoxyl acetate [170.5 (s, C1), 72.0 (t, C2), 59.1 (s, OCH_3)], Val 1 [174.1 (s, CO), 54.7 (d, C2), 30.8 (d, C3)], MeLeu 1 [172.9 (s, CO), 56.4 (d, C2), 38.1 (t, C3), 25.9 (d, C4)], Thr [169.2 (s, CO), 71.0 (d, C3), 53.7 (d, C2), 17.8 (q, C4)], MeβAla [172.7 (s, CO), 45.3 (t, C3), 34.7 (q, NCH_3), 34.7 (t, C2)], Val 2 [174.1 (s, CO), 55.5 (d, C2), 31.8 (d, C3)], MeVal [61.0 (d, C2), 31.5 (q, NCH_3), 31.0 (d, C3)], β-Ala [173.0 (s, CO), 35.7 (t, C2), 35.5 (t, C3)], Ile [175.9 (s, CO), 53.7 (d, C2), 38.8 (d, C3), 15.4 (q, C6)], MeIle [172.5 (s, CO), 55.5(d, C2), 33.8 (d, C3), 31.7 (q, NCH_3), 15.8 (q, C6)], MeAla [171.2 (s, CO), 57.5 (d, C2), 29.7 (q, NCH_3), 15.0 (q, C3)], Ala [174.6 (s, CO), 52.2 (d, C2), 17.6 (q, C3)], Leu [174.8 (s, CO), 48.8 (d, C2), 40.7 (t, C3), 25.5 (d, C4)], MeLeu [172.4 (s, CO), 63.0 (d, C2), 38.5 (t, C3), 38.3 (q, NCH_3), 25.6 (d, C4)].

Amino Acid Analysis by GC/MS The *N*-trifluoroacetamide-*n*-butyl esters of the hydrolysate were prepared and analysed as described in Section 7.4.1.

Methanolysis of 58 Theonellapeptolide IIId (**58**) (11.5 mg) was dissolved in dry MeOH (2.4 mL) with NaOMe/MeOH (4%) (0.7 mL). The methanolysis product **59** (9.8 mg) was isolated and purified as described in Section 7.4.1.

Methyl Ester of Theonellapeptolide IIId (59) FABMS, *m / z* 1437 $(M+H)^+$, 1093, 1051, 1037, 1008, 881, 853, 726, 711, 697, 669, 654, 603, 581, 556, 521, 429. 400, 344, 337, 299, 223, 172, 100, 72, 57, 45.

Determination of Absolute Configuration The configuration of the component amino acids of theonellapeptolide IIId was determined as described in Section 7.4.4.

References

1. Cragg, G. M.; Newman, D. J.; Snader, K. M. *J. Nat. Prod.* **1992,** 60, 52-60.
2. Seltzer, R. J. *Science* **1975**, 53 (1), 13-14.
3. de Vries, D. J.; Hal, M. *Drug Dev. Res.* **1994**, 33, 161-173.
4. McConnell, O.; Longley, R. E.; Koehn, F. E. *The Discovery of Natural Products with Therapeutic Potential.* Boston, MA, Butterworth-Heinemann, **1994**, PP 109-174.
5. Bergquist, P. R. *Colloques internationaux du C. N. R. S.* **1979**, 291, 383-392.
6. Garson, M. J.; Murphy, P. T. *Chemistry in Australia* **1989**, (10), 358-361.
7. Carte, B. K. *Current Opinion in Biotechnology* **1993**, 4, 275-279.
8. Ruggieri, G. D. *Science*, **1976**, 194, 491-497.
9. Okuda, R. K.; Klein, D.; Kinnel, R. B.; Li, M.; Scheuer, P. J. *Pure & Appl. Chem.* **1982**, 54 (10), 1907-1914.
10. Younken, H. W. (Ed.) *Food-Drugs from the Sea*: Proceedings **1969**. Washington DC, Marine Technology Society, 1970.
11. Kau, P. N. *Impact of Science on Society* **1979**, 29 (2), 123-134.
12. Reed, J. K.; Pomponi, S. A. *American Academy of Underwater Sciences Ninth Annual Scientific Diving Symposium*, Woods Hole, Massachusetts, **1989**, pp273-287.
13. Rinehart, K. L.; Shield, L. S. *Proc. Symp.* Dedicated to Late Prof. H. Umezawa, Antibiotic Research Associates, Japan. **1987**, pp194-212.
14. Baker, J. T. *Hydrobiologia*, **1984**, 116/117, 29-40.
15. Munro, M. H. G.; Blunt, J. W.; Lake, R. J.; Litaudon, M.; Battershill, C. N.; Page, M. *Sponges, in Time and Space* (van

Soest, van Kempen, Braekman, eds.), Balkema, Rotterdan, **1994**, pp473-484.

16. Faulkner, D. J. *Chemistry in Britain* **1995**, (9), 680-684.
17. *MarinLit*, **1997**. A marine literature database, maintained by the Marine Chemistry Group, University of Canterbury, Christchurch, New Zealand.
18. de Vries, D. J.; Beart, P. M. *Trends in Pharmacological Sciences* **1995**, 16(8), 275-279.
19. Palmer, M. *Natural products: Rapid Utilization of Sources for Drug Discovery and Development* Mulford, N. (Ed.) International Business Communications. U.S.A. **1996**, pp1.95-1.109.
20. Faulkner, D. J.; Unson, M. D.; Bewley, C. A. *Pure & Appl. Chem.* **1994**, 66, 1983-1990.
21. Konig, G. M.; Wright, A. D. *Planta Med.* **1996**, 62, 193-211.
22. Scheuer, P. J.; Ed.; *Bioorganic Marine Chemistry*, Springer-Verlag: Heiderberg, **1987-1992**, Vols. 1-6.
23. Faulkner, D. J. *J. Nat. Prod. Rep.* **1992,** 9, 323-364.
24. Shimizu, Y. *Progress in the Chemistry of Organic Natural Products*; Herz, W.; Grisebach, H.; Kirby, G. W.; Tamm, C., Eds.; Springer-Verlag: Wien, **1984**, 45, pp235-264.
25. Yotsu, M.; Yamazaki, T.; Meguro, Y.; Endo, A.; Murata, M.; Naoki, H.; Yasumoto, T. *Toxicon* **1987**, 25, 225-228.
26. Yasumoto, T.; Yasumura, D.; Yotsu, M; Michishita, T.; Endo, A.; Kotaki, Y. *Agric. Biol. Chem.* **1986**, 50, 793-795.
27. Noguchi, T.; Jeon, J. K.; Arakawa, O.; Sugita, H.; Deguchi, Y.; Shida, Y.; Hashimoto, K. *J. Biochem.* **1986**,99, 311-314.

28 Ciavatta, M. L.; Trivellone, E.; Cimino, G.; Uriz, M. J. *Tetrahedron Lett.* **1994**, 35, 7871-7874.

29. Hamel, E.; Blokhin, A. V.; Nagle, D. G.; Oyo, H. D.; Gerwick, W. H. *Drug Development Research* **1995**, 34, 110-120.
30. Rutzler, K. *New Perspectives in Sponge Biology* (Rutzler, K., Ed.), Smithsonian Institution Press, Washington DC, **1990**, pp455-466.
31. Stierle, A. C.; Cardellina, J. H. II; Singleton, F. L. *Experientia* **1988**, 44, 1021
32. Proksch, P. *Toxicon* **1994**, 32, 639-655.
33. Gil-Turnes, M. S.; Hay, M. E.; Fenical, W. *Science*, **1989**, 246, 116-118
34 Gil-Turnes, M. S.; Fenical, W. *Biol. Bull.* **1992**, 182, 105-108.
35. Bakus, G. J.; Targett, N. M.; Schulte, B. *J. Chem. Ecol.* **1986**, 12, 951-987.
36. Neeman, I.; Fishelson, L.; Kashman, Y. *Mar. Biol.* **1975**, 30, 293-296.
37. Quinn, J. F. *Oecologia* **1982**, 54, 129-135.
38. Benayahu, Y.; Loya, Y. *Bull. Mar. Sci.* **1981**, 31, 514-522.
39. Joffe, S.; Thomas, R. *AgBiotech. News Inform.* **1989**, 1, 697-700.
40. Seltzer, R. J. *Science* **1975**, 53, 13-14.
41. Rinehart, K. L.; Gloer, J. B.; Cook, J. C.; Mizsak, S. A. & Scahill, T. A. *J. Am. Chem. Soc.* **1981**, 103, 1857-1859.
42. Jayson, G. C.; Crowther, D.; Prendiville, J.; McGown, A. T.; Scheild, C.; Stern, P.; Young, R.; Brenchley, P.; Chang, G.; Owens, S.; Pettit, G. R. *Br. J. Cancer* **1995**, 72, 461-468.
43. Pettit, G. R.; Kamano, Y.; Herald, C. L.; Tuiman, A. A.; Boettner, F. E.; Kizu, H.; Schmidt, J. M.; Baczynskyi, L.; Tomer, K. B.; Bontems, R. J. *J. Am. Chem. Soc.* **1987**, 109, 6883-6885.

44. Rinehart, K. L.; Holt, T. G.; Fregeau, N. L.; Keifer, P. A.; Wilson, G. R.; Perun, T. J.; Sakai, R.; Thomson, A, G.; Stroh, J. G.; Shield, L. S.; Seigler, D. S.; Li, L. H.; Martin, D. G., Grimmelikhuijzen, C. P. J.; Gade, G. *J. Nat. Prod,* **1990**, 53, 771-792.

45. Fuller, R. W.; Cardellina, J. H.; Jurek, J. H.; Scheuer, P. J.; Alvarado-Lindner, B.; Mcguire, M.; Gray, G. N.; Steiner, J. R. Clardy, J. *J. Med. Chem.* **1994**, 37, 4407-4411.

46. Uemura, D.; Takahashi, K.; Yamamoto, T.; Katayama, C.; Tanaka, J.; Okumura, Y.; Hirata, Y. *J. Am. Chem. Soc.* **1985**, 107, 4796-4798.

47. Litaudon, M.; Hart, J. B.; Blunt, J. W.; Lake, R. J.; Munro, M. H. G. *Tetrahedron Lett.* **1994**, 35, 9435-9438.

48. Pettit, G. R.; Cichacz, Z. A.; Gao, F.; Herald, C. L.; Boyd, M. R.; Schmidt, J. M.; Hooper, J. N. A. *J. Org. Chem.* **1993**, 58, 1302-1304.

49. De Freitas, J. C.; Blankemeier, L. A.; Jacobs, R. S. *Experientia* **1984**, 40, 864-865.

50. Wheeler, L. A.; Sachs, G. De Freitas, J. C.; Goodrum, D.; Woldemussie, E.; Muallem, S. *J. Biol. Chem.* **1987**, 262, 6531-6538.

51. Trischman, J. A.; Tapiolas, D. M.; Jensen, P. R.; Dwight, R.; Fenical, W. *J. Am. Chem. Soc.* **1994**, 116, 57-758.

52. Gunasekera, S. P.; Cranick, S.; Longley, R. E. *J. Nat Prod.* **1989**, 52, 757-761.

53. Needham, J.; Kelly, M. T.; Ishegi, M.; Andersen, R. J. *J. Org. Chem.* **1994**, 59, 2058-2063.

54. Stratmann, K.; Burgoyne, D. L.; Moore, R. E.; Patterson G. M. L.; Smith, C. D. *J. Org. Chem.* **1994**, 59, 7219-7226.

55. Quinoa, E.; Adamczeski, M.; Crews, L. R.; Bakus, G. J. *J. Org. Chem.* **1986**, 51, 4994-4997.

56. Fusetani, N.; Matsunaga, S.; Matsumoto, H.; Takebayashi, Y. *J. Am. Chem. Soc.* **1990**, 112, 7053-7054.

57. Moore, R. E; Bornemann, V.; Niemczura, W. P.; Gregson, J. M.; Chen. J. L.; Norton, T. R.; Patterson G. M. L.; Helms, G. L. *J. Am. Chem. Soc.* **1989**, 111, 6128-6132.

58. Nagai, H.; Murata, M.; Torigoe, K.; Satake, M.; Yasumoto, T. *J. Org. Chem.* **1992**, 57, 5448-5453.

59. Larson, E. *et al. Bioactive Compounds from the Sea*, Humm, H. J.; Lane, C. E. Eds, Marcel Decker, New York, **1974**, pp139-141.

60. Kao, C. Y. *Bioactive Compounds from the Sea*, Humm, H. J.; Lane, C. E. Eds, Marcel Decker, New York, **1974**, pp115-121.

61. Narahashi, T.; Haas, H. G.; Therrien, E. F. *Science* **1967**, 157, 1441-1442.

62. Jacobs, R. S.; Culver, P.; Langdon, R.; O'Brien, T.; White, S. *Tetrahedron* **1985**, 41, 981-984.

63. Moore, R. E.; Scheuer, P. J. *Science* **1971**, 172, 495-498.

64. Moore, R. E.; Woolard, T. X.; Bartolini, G. *J. Am. Chem. Soc.* **1980**, 102, 7370.

65. Tachibana, K.; Scheuer, P. J.; Tsukitani, Y.; Kikuchi, H.; Van Engen, D.; Clardy, J.; Gopichand, Y.; Schmitz, F. J. *J. Am Chem. Soc.* **1981**,103, 2491-2494.

66. Murakami, K.; Oshima, Y.; Yasumoto, T. *Bull. Jap. Soc. Sci. Fish* **1982**, 48, 69-72.

67. Barns, G.; Blunt, J. W.; Munro, M. H. G.; Perry, N. B.; Lake, R. G.; Haystead T. A. *Nat. Prod. Lett.* In Press.

68. Cohen, P.; Cohen, P. T. W. *J. Biol. Chem.* **1989**, 264, 21435-21438

69. Borowitzka, L. J. *Hydrobiologia*, **1981**, 81, 33-46.

70. Ben-Amotz, A. *Marine Biotechnology: Pharmaceutical and Bioactive Natural Products*, Attaway, D. H.; Zaborsky, O. R. (Ed.) **1993, Vol**. 1, pp411-417.

71. Look, S. A.; Fenical, W.; Matsumoto, G. K.; Clardy, J. *J. Org. Chem.* **1986**, 51, 5140-5145.

72. Bongiorni, L.; Pietra, F. *Chemistry & Industry* **1996**, (1), 54-58.

73. Breivik, H.; Boerretzen, B.; Joergensen, T. E. *PCT. Int. Appl. WO* 8703, 899.

74. McGovern, P. E.; Michel, R. H. *Acc, Chem, Res.*, **1990**, 23, 152-158.

75. Chapman, V. J. *Marine Chem. Rev.* **1968**, 3-13.

76. Bakuni, D. S., Silva, M. *Bot. Mar.* **1974**, 17, 40-51.

77. Colwell, R. R. *Oceanus* **1984**, 27, 3-12.

78. Bruening, R. C.; Oltz, E. M.; Furukawa, J.; Nakanishi, K.; Kustin, K. *J. Nat. Prod.* **1986**, 49, 193-204.

79. Cragg, G. M.; Newman, D. J.; Weiss, R. B. *Seminars in Oncology*, **1997**, 24, 156-163.

80. Bailey, P. D. *An Introduction to Peptide Chemistry,* John Wiley & Sons, West Sussex, England, **1990**, pp48-51.

81. Determann, H. *Gel Chromatography*, Springer-Verlag, Berlin, **1969**, pp1-5.

82. Henke, H. *Preparative Gel Chromatography on Sephadex LH-20*, Huthig Verlag Heidelberg, Germany, **1995**, pp4-32.
83. Mant, C. T.; Hodges, R. S., in Gooding, K.; Regnier, F. E. (Editors), *HPLC of Biological Macromolecules: Methods and Applications*, Marcel Dekker, New York, **1990**, pp 301.
84. Zhou, N. E., Mant, C. T.; Kirkland, J. J., Hodges, R. S. *J. Chromatogr.* **1991**, 548, 179-193.
85. River, J.; McClintock, R.; Galyean, R.; Anderson, H. *J. Chromatogr.* **1984**, 288, 303-328.
86. Shenbagamurthi, P.; Naider, F.; Beker, J. M.; Steinfeld, A. S. *J. Chromatogr.* **1983**, 256, 117-125.
87. Keutmann, H. T.; Potts, J. T. Jr. Anal. *Biochem.*, **1969**, 29, 185-185.
88. Moore, S.; Spackman, D. H.; Stein, W. H. *Anal. Chem.* **1958**, 30, 1185-1190.
89. Moore, S.; Stein, W. H. *J. Biol. Chem.* **1951**, 192, 663-681.
90. Moore, S.; Stein, W. H *J. Biol. Chem.* **1954**, 211, 895-906.
91. Stein, S.; Bohlen, P.; Stone, J.; Dairman, W.; Udenfriend, S. *Arch. Biochem. Biophys.* **1973**, 155, 202.
92. Lange, H. W.; Lustenberg, N.; Hempel, K. Fresenius'z. *Anal. Chem.* **1972**, 261, 337-343.
93. Lustenberg, N.; Lange, H. W.; Hempel, K. *Angew. Chemie, Int. Ed. Engl.* **1972**, 11, 227-229.
94. Husek, P.; Macek, K. *J. Chromatogr.* **1975**, 113, 139-230.
95. Gamerith, G. *J. Chromatogr.* **1983**, 256, 326-330.
96. Zumwalt, R. W.; Desgres, J.; Kuo, K. C.; Pautz, J. E.; Gehrke, C. W. *J. Assoc. Off. Anal Chem.* **1987**, 70, 252-262.
97. Cohen, S. A.; Strydom, D. J. *Anal. Biochem.* **1988**, 174, 1-16.

98. Bruton, C. J.; Hartley, B. S. *J. Mol. Biol.* **1970**, 52, 165- 178.

99. Edman, P *Acta Chem. Scand.* **1950**, 4, 283-293.

100. Needleman, S. B.(Ed.) *Protein Sequence Determination*, Chapman & Hall, **1970**, pp141-165.

101. Eckart, K. *Mass Spectrom. Rev.* **1994**, 13, 23-55.

102. Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, 14, 49-73.

103. Cililla, F. J.; Yadav, S. P.; Brew, K.; Mendez, E. *J. Chromatogr.* **1991**, 548, 303-310.

104. Manning, J. M.; Moore, S. *J. Biol Chem.* **1968**, 243, 5591-5597.

105. Hasegawa, M.; Matsubara, I. *Anal. Biochem.* **1975**, 63, 308-320.

106. Gil-Av, E. *J. Mol. Evol.* **1975**, 6, 131-144.

107. Abe, I.; Fujimoto, N.; Nakahara, T. *Chem. Lett.* **1995**, 113-114.

108. LePage, J. N.; Lindner, W.; Davies, G.; Seitz, D. E.; Karger, B. L. *Anal. Chem.* **1979**, 51, 433-435.

109. Marfey, P. *Carlsberg Res. Commun.* **1984**, 49, 591-596.

110. Suwan, S.; Isobe, M.; Ohtani, I.; Agata, N.; Mori, M.; Ohta, M. *J. Chem. Soc. Perkin Trans.* **1995**, (1), 765-775.

111. Ireland, C. M.; Molinski, T. F.; Roll, D. M.; Zabriskie, T. M.; Mckee, T. C.; Swersey, J. C.; Foster, M. P. Natural Product Peptides from Marine organisms. *Bioorganic Marine Chemistry*, Scheuer, P. J. Ed. Springer-Verlag, Heideberg, **1989, Vol.3**, pp1-46.

112. Matsunaga, S.; Fusetani, N.; Konosu, S. *J. Nat. Prod.* **1985**, 48, 236-241.

113. Brabley, S.; Hammann, P.; Kluge, H.; Wink, J.; Kricke, P. Zeeck, A. *J. Antibiotics* **1991**, 44, 797-800.

114. Noltemeyer, M.; Sheldrick, G. M.; Hoppe, M. V.; Zeeck, A. *J. Antibiotics* **1982**, 35, 549-555.

115. Cardellina II, J. H.; Munro, M. H. G.; Fuller, R. W.; Manfredi, K. P.; Mckee, T. C.; Tischler, M.; Bokesch, H. R.; Gustafson, K. R.; Beutler, G. A.; Boyd, M. R. *J. Nat. Prod.* **1993**, 56 (7), 1123-1129.

116. Blunt, J. W.; Dumdei, E. J.; Munro, M. H. G.; Stirling, D. J.; Li, S. *Eighth International Symposium on Marine Natural Products* (Canary Islands), **1995**, P2, 114-115.

117. Northcote, P. T.; Blunt, J. W.; Munro, M. H. G. *Tetrahedron Lett.*, **1991**, 32, 6411-6414.

118. Dumdei, E. J.; Blunt, J. W.; Munro, M. H. G.; Pannell, L. K. *J. Org. Chem* **1997**, 62, 2636-2639.

119 Matsunaga, S.; Fujiki, H.; Sakata, D; Fusetani, N. *Tetrahedron* **1991**, 47, 2999-3006.

120. Beutler, J. A.; McKee, T. C.; Fuller, R. W.; Tischler, M.; Cardellina II, J. H.; Snader, K. M.; McCloud, T.; Boyd, M. R. *Antiviral Chemistry & Chemotherapy* **1993**, 4(3), 167-172.

121. Albano, R. M.; Mourao, P. A. S. *Biochim. Biophys Acta* **1983**, 760, 192-196.

122. Albano, R. M.; Mourao, P. A. S. *J. Biol. Chem.* **1986**, 261, 758-765.

123. Stirling, D. J. *Studies on Marine Natural Products* (Ph. D. thesis) University of Canterbury, 1996, pp 178-179.

124. Perry, N. B.; Blunt W. J.; Munro, M. H. G.; Pannell L. K. *J. Am. Chem. Soc.* **1988**, 110, 4850-4851.

125. Toth, S. I.; Schmitz, F. J. *J. Nat. Prod.* **1994**, 57, 123-127.

126. Kobayashi, M.; Higuchi, K.; Murakami, N.; Tajima, H.; Aoki, S. *Tetrahedron Lett.*, **1997**, 38, 1859-2862.
127. Kusumi, T.; Ohtami, I.; Kakisawa, H. *Tetrahedron Lett.* **1988**, 29, 4731-4734.
128. Mamamoto, T.; Seto, H.; Beppu, T. *J. Antibiotics*, **1983**, 36, 646-650.
129. Komiyama, K.; Okada, H.; Tomisaka, S.; Miyano, T.; Funayama, S.; Umwzawa, I. *J. Antibiotics*, **1985**, 38, 220-223.
130. Hayahawa, Y.; Sohda, K.; Seto, H. *J. Antibiotics*, **1996**, 49, 980-984.
131. Davies-Coleman, M. T.; Faulkner D. J.; Dubowchik, G. M.; Roth, G. P.; Polson C.; Fairchild, C. *J. Org. Chem.* **1993**, 58, 5925-5930.
132. Uno, M.; Ohta, S.; Ohta, E.; Ikegami, S. *J. Nat. Prod.* **1996**, 59, 1146-1148.
133. Fukami, A.; Ikeda, Y.; Kondo, S.; Naganawa, H.; Takeuchi, T.; Furuya, S. *Tetrahedron Lett.* **1997**, 38, 1201-1202.
134. Wang, G.; Kuramoto, M.; Uemura, D.; Yamada, A.; Yamaguchi, K.; Yazawa, K. *Tetrahedron Lett.* **1996**, 37, 1813-1816.
135. Quinoa, E.; Crews, P. *Tetrahedron Lett.* **1987**, 28, 2467-2468.
136. Kobayashi, M.; Lee, N. K.; Shibuya, H.; Momose, T.; Kitagawa, I. *Chem. Pharm. Bull.* **1991**, *39*, 1177-1184.
137. Kitagawa, I.; Kobayashi, M.; Lee, N. K.; Shibuya, H.; Kawata, Y.; Sakiyama, F. *Chem. Pharm. Bull.* **1986**, *34*, 2664-2667.
138. Kitagawa, I.; Lee, N. K.; Kobayashi, M.; Shibuya, H. *Chem. Pharm. Bull.* **1987**, *35*, 2129-2132.
139. Kobayashi, M.; Kanzaki, K.; Katayama, S.; Ohashi, K.; Okada, H.; Ikegami, S.; Kitagawa, I. *Chem. Pharm. Bull.* **1994**, *42* 1410-1415.

140. Kitagawa, I.; Lee, N. K.; Kobayashi, M.; Shibuya, H. *Tetrahedron* **1991**, *47*, 2169-2180.

141. Kitagawa, I.; Ohashi, K.; Kawanishi, H.; Shibuya, H.; Shinkai, K.; Akedo, H. *Chem. Pharm. Bull.* **1989**, *37*, 1679-1681.

142. Perry, N. B.; Blunt, J. W.; Munro, M. H. G.; Thompson, A. M. *J Org. Chem.* **1990**, *55*, 223-227.

143. Leimer, K. R.; Rice, R. H.; Gehrke, C. W. *J. Chromatogr.* **1977**, *141*, 121-144

144. Lawless, J. G.; Chadha, M. S. *Anal. Biochem.* **1971**, *44*, 473-485.

145. Gelpi, E.; Koenig, W. A.; Gilbert, J.; Oro, J. *J. Chromatogr. Sci.* **1969**, 7, 604-613.

146. Roepstorff, P; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

147. Mueler, D. R.; Eckersley, M.; Rochter, W. J. *Org. Mass Spectrom.* **1988**, 23, 217-222.

148. Hemling, M. E. *Pharmaceutical Research* **1987**, 4(1), 5-15.

149. Clench, M. r.; Garner, G. V.; Gorden, D. B.; Barber, M. *Biomed. Mass Spectrom.* **1985**, 12, 355-357.

150. Naylor, S.; Findeis, A. F.; Gibson, B. W.; Williams, D. H. *J. Am. Chem. Soc.* **1986**, 108, 6359-6363.

151. Naylor, S.; Moneti, G.; Guyan, S. *Biomed. Environ. Mass Spectrom.*. **1988**, 17, 393-397.

152. Biemann, K.; Martin, S. A. *Mass Spectrom. Rev.* **1987**, 6, 1-76.

153. Hinds, M. G.; Norton, R. S. *Methods in Molecular Biology*, Humana Pree, NJ, **1994**, Vol. 36, pp131-154.

154. Sheldrick, G. M. *J. Appl. Crystallogr.* In Press

155. Bernardinelli, G.; Jefford, C. W.; Sakai, R.; Higa, T. *Proceedings, Seventh International Symposium on Marine Natural Products,* July, 1992, Capri, Italy. P29.

156. Van Der Haar, J.; Kip, J.; Kraak, J. C. *J. Chromatogr.* **1988**, *445*, 219-224.

157. Armani, E.; Barazzoni, L.; Dossena A.; Marchelli, R. *J. Chromatogr.* **1988**, *441*, 287-298.

158. Takeuchi, T.; Asai, H.; Hashimoto, Y.; Watanabe, K.; Ishii, D. *J. Chromatogr.* **1985**, *331*, 99-107.

159. Weinstein, S.; Weiner, S. *J. Chromatogr.* **1984**, *303*, 244-250.

160. Nimura, N.; Toyama, A.; Kinoshita, T. *J. Chromatogr.* **1984**, *316*, 547-552.

161. Lindner, W.; Lepage, J. N.; Davies, G.; Seitz, D. E.; Karger, B. *J. Chromatogr.* **1979**, *185*, 323-343.

162. Evans, M. B.; Dale, A. D.; Little, C. J. *Chromatographia* **1980**, 13, 5-10.

Appendix One

X-Ray Data for Crystal Structure

Table A1.1 Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$ x 10^3) for **50**. U(eq) is defined as one third of the trace of the orthogonalized U_{ij} tensor.

	x	y	z	U(eq)
O11	1685(6)	1091(3)	1629(2)	55(2)
C11	2377(9)	760(5)	1508(2)	42(2)
C21	3037(7)	217(4)	1684(2)	43(2)
C3'1	2506(8)	7(5)	1999(2)	46(2)
C31	3114(10)	-535(7)	2201(3)	73(3)
C41	4170(12)	-270(10)	2327(4)	126(6)
C51	2396(13)	-804(9)	2455(3)	116(6)
N1	3206(6)	-376(3)	1465(2)	38(2)
CN1	4346(8)	-548(6)	1394(3)	72(3)
O22	1470(5)	-395(3)	1345(1)	42(2)
C12	2367(7)	-642(4)	1312(2)	28(2)
C22	2532(6)	-1285(4)	1116(2)	32(2)
C3'2	2454(8)	-1914(5)	1347(2)	46(2)
C32	2796(9)	-2585(5)	1195(2)	53(3)
C42	4009(10)	-2648(6)	1165(3)	93(4)
C52	2339(12)	-3201(5)	1391(3)	92(4)
N2	1719(5)	-1353(3)	864(2)	34(2)
O33	2691(5)	-884(3)	463(1)	41(2)
C13	1854(7)	-1174(4)	561(2)	32(2)
C23	939(6)	-1304(4)	331(2)	31(2)
C33	229(7)	-687(4)	287(2)	46(2)
N3	315(6)	-1915(4)	428(2)	33(2)
O44	1699(5)	-2610(3)	374(1)	44(2)
C14	754(8)	-2508(5)	440(2)	35(2)
C24	98(7)	-3117(4)	579(2)	44(2)
C34	515(9)	-3806(5)	445(3)	81(4)
N4	-1064(6)	-3039(4)	529(2)	44(2)
CN4	-1424(8)	-3127(5)	193(2)	59(3)
O55	-2769(6)	-3093(3)	701(2)	55(2)
C15	-1803(8)	-3029(4)	762(2)	44(2)
C25	-1473(8)	-2929(4)	1121(2)	45(2)
C35	-1827(10)	-3531(5)	1333(3)	69(3)
C45	-1338(11)	-4188(5)	1196(3)	76(4)
C55	-1405(19)	-3400(7)	1691(3)	145(8)
C65	-1801(17)	-3941(10)	1927(4)	152(7)
N5	-1921(6)	-2258(4)	1219(2)	53(2)
CN5	-3069(9)	-2191(7)	1273(4)	117(6)
O66	-271(5)	-1793(3)	1167(1)	44(2)
C16	-1226(9)	-1716(5)	1234(2)	44(2)
C26	-1648(9)	-1042(5)	1364(2)	58(3)
C36	-1676(13)	-1076(6)	1738(3)	88(4)
C46	-535(14)	-1108(8)	1889(3)	99(5)
C56	-2376(18)	-437(7)	1862(4)	133(8)
C66	-2931(21)	-458(10)	2087(10)	330(28)
N6	-975(6)	-494(4)	1249(2)	41(2)
O77	-2366(6)	142(4)	1063(2)	70(2)
C17	-1403(9)	66(5)	1098(2)	54(3)

C27	-611(8)	563(5)	958(2)	53(3)
C37	-930(9)	1318(5)	999(3)	62(3)
N7	-574(8)	1612(6)	1304(3)	88(4)
O88	-1819(9)	1267(7)	1600(2)	133(5)
C18	-1017(10)	1615(6)	1570(3)	64(3)
C28	-531(8)	2009(5)	1851(2)	52(3)
C38	-1038(11)	1839(6)	2171(3)	75(4)
C48	-622(16)	1077(6)	2269(3)	118(6)
C58	-737(11)	2341(6)	2445(3)	81(4)
C68	-1310(16)	2219(8)	2755(3)	121(6)
N8	-544(6)	2743(4)	1765(2)	49(2)
CN8	-1614(10)	3056(8)	1728(3)	94(4)
O99	1229(6)	2800(3)	1737(2)	57(2)
C19	360(9)	3083(5)	1705(2)	42(2)
C29	365(9)	3830(5)	1604(2)	56(3)
C39	725(20)	4280(6)	1893(3)	133(8)
C49	919(20)	5048(6)	1767(3)	151(8)
C59	238(23)	4240(8)	2155(4)	186(12)
N9	1130(6)	3932(4)	1345(2)	51(2)
O1010	-74(6)	4178(4)	962(2)	67(2)
C110	866(8)	4111(5)	1053(2)	48(3)
C210	1770(8)	4243(5)	821(3)	63(3)
C310	1978(9)	3671(5)	577(3)	66(3)
N10	2593(6)	3110(4)	717(2)	53(2)
CN10	3758(8)	3204(6)	707(3)	69(3)
O1111	1094(6)	2526(3)	861(2)	69(2)
C111	2077(10)	2572(5)	851(2)	52(3)
C211	2776(8)	2039(4)	1022(2)	49(3)
C311	2229(8)	1332(4)	1004(2)	52(3)
C411	2255(11)	1018(6)	664(3)	80(4)
N11	2933(6)	2282(3)	1351(2)	40(2)
O11	2799(5)	869(3)	1215(2)	48(2)
O1212	4741(6)	2156(4)	1370(2)	71(2)
C112	3851(9)	2354(5)	1492(2)	48(2)
C212	3833(7)	2698(4)	1823(2)	40(2)
C3'12	4050(8)	2189(5)	2093(2)	49(3)
C312	3897(10)	2475(6)	2429(2)	68(3)
C412	4427(12)	2010(9)	2686(3)	111(5)
C512	2718(12)	2568(8)	2508(3)	104(5)
N12	4599(5)	3284(4)	1819(2)	38(2)
CN12	5714(7)	3149(5)	1892(2)	47(2)
O1313	3337(6)	3897(3)	1559(2)	52(2)
C113	4245(8)	3865(5)	1677(2)	38(2)
C213	5000(7)	4474(4)	1636(2)	38(2)
C313	4429(8)	5148(5)	1638(2)	54(3)
C413	5176(9)	5743(5)	1569(3)	77(3)
C513	3885(10)	5251(6)	1972(3)	74(3)
N13	5597(7)	4349(4)	1341(2)	45(2)
O1414	7260(5)	4436(3)	1567(2)	55(2)
C114	6679(9)	4357(5)	1323(2)	47(2)
C214	7149(8)	4227(5)	996(2)	53(3)
O1514	6533(6)	3779(4)	797(2)	68(2)
C014	6670(12)	3060(6)	885(3)	88(4)
X115	2596(5)	-25(3)	85(2)	61(2)
X215	-3712(6)	-4359(4)	566(2)	70(2)
X315	-4710(6)	-2647(4)	374(2)	83(2)
X415	-612(6)	4349(4)	331(2)	69(2)

X515	2964(6)	5198(4)	248(2)	87(2)
X615	6108(6)	3975(4)	144(2)	76(2)
X715	3417(6)	-3416(4)	248(2)	83(2)
X815	-1979(6)	-1534(4)	466(2)	75(2)
X915	4452(7)	4941(4)	772(2)	91(3)
X1015	-962(7)	792(4)	109(2)	84(2)
X1115	11017(8)	4768(5)	-51(2)	103(3)
X1215	4788(9)	-1336(6)	585(3)	122(3)
X1315	7888(14)	1938(9)	260(4)	193(6)
X1415	6228(15)	2599(10)	-13(5)	207(7)
X1515	6046(16)	1289(10)	945(5)	216(7)
X1615	9383(18)	3118(11)	-59(5)	241(8)
X1715	-3603(19)	-1015(12)	922(5)	249(9)
X1815	-5265(104)	931(55)	888(25)	1058(97)
X1915	-3538(8)	-501(5)	574(2)	111(3)
X2015	-4158(11)	1024(7)	1719(3)	151(5)
X2115	6563(34)	1427(22)	368(9)	416(20)

Table A1.2 Bond lengths[Å] for **50**.

	Bond lengths		**Bond lengths**
O11-C11	1.192(11)	C28-N8	1.480(13)
C11-O11	1.343(11)	C28-C38	1.51(2)
C11-C21	1.530(14)	C38-C58	1.55(2)
C21-N1	1.488(11)	C38-C48	1.63(2)
C21-C3'1	1.521(13)	C58-C68	1.49(2)
C3'1-C31	1.55(2)	N8-C19	1.335(12)
C31-C51	1.48(2)	N8-CN8	1.480(14)
C31-C41	1.51(2)	O99-C19	1.227(11)
N1-C12	1.331(11)	C19-C29	1.522(14)
N1-CN1	1.495(12)	C29-N9	1.449(13)
O22-C12	1.229(10)	C29-C39	1.55(2)
C12-C22	1.513(12)	C39-C59	1.25(2)
C22-N2	1.462(10)	C39-C49	1.61(2)
C22-C3'2	1.563(12)	N9-C110	1.304(12)
C3'2-C32	1.518(13)	O1010-C110	1.241(11)
C32-C42	1.53(2)	C110-C210	1.506(14)
C32-C52	1.56(2)	C210-C310	1.53(2)
N2-C13	1.313(11)	C310-N10	1.459(13)
O33-C13	1.258(10)	N10-C111	1.353(13)
C13-C23	1.510(12)	N10-CN10	1.469(13)
C23-N3	1.484(10)	O1111-C111	1.234(12)
C23-C33	1.510(12)	C111-C211	1.54(2)
N3-C14	1.285(11)	C211-N11	1.456(12)
O44-C14	1.229(10)	C211-C311	1.546(13)
C14-C24	1.557(13)	C311-O11	1.447(11)
C24-N4	1.476(12)	C311-C411	1.53(2)
C24-C34	1.548(13)	N11-C112	1.295(12)
N4-C15	1.340(12)	O1212-C112	1.283(12)
N4-CN4	1.472(12)	C112-C212	1.528(13)
O55-C15	1.241(11)	C212-N12	1.494(11)
C15-C25	1.552(14)	C212-C3'12	1.523(13)
C25-N5	1.486(12)	C3'12-C312	1.511(14)
C25-C35	1.536(13)	C312-C512	1.52(2)
C35-C45	1.53(2)	C312-C412	1.55(2)
C35-C55	1.59(2)	N12-C113	1.355(11)
C55-C65	1.52(2)	N12-CN12	1.451(11)
N5-C16	1.374(12)	O1313-C113	1.238(10)
N5-CN5	1.459(13)	C113-C213	1.530(13)
O66-C16	1.236(11)	C213-N13	1.452(12)
C16-C26	1.520(14)	C213-C313	1.500(13)
C26-N6	1.444(11)	C313-C413	1.520(14)
C26-C36	1.55(2)	C313-C513	1.55(2)
C36-C46	1.56(2)	N13-C114	1.355(12)
C36-C56	1.61(2)	O1414-C114	1.255(11)
C56-C66	1.17(3)	C114-C214	1.496(14)
N6-C17	1.370(13)	C214-O1514	1.430(12)
O77-C17	1.222(12)	O1514-C014	1.464(13)
C17-C27	1.505(14)	X1515-X1815#1	1.80(12)
C27-C37	1.541(14)	X1715-X1915	1.76(2)
C37-N7	1.455(14)	X1815-X1515#2	1.80(12)
N7-C18	1.235(14)		
O88-C18	1.219(13)		
C18-C28	1.52(2)		

Table A1.3 Bond angles [°] for **50**

	Bond angles		**Bond angles**
O11-C11-O11	125.4(9)	O66-C16-N5	120.5(8)
O11-C11-C21	124.6(9)	O66-C16-C26	121.5(8)
O11-C11-C21	109.3(9)	N5-C16-C26	117.8(8)
N1-C21-C3'1	112.0(7)	N6-C26-C16	109.0(7)
N1-C21-C11	109.0(7)	N6-C26-C36	111.9(8)
C3'1-C21-C11	111.1(8)	C16-C26-C36	108.9(9)
C21-C3'1-C31	115.6(8)	C26-C36-C46	112.4(10)
C51-C31-C41	113.9(11)	C26-C36-C56	107.3(12)
C51-C31-C3'1	109.2(11)	C46-C36-C56	113.6(12)
C41-C31-C3'1	112.3(11)	C66-C56-C36	124(2)
C12-N1-C21	118.9(7)	C17-N6-C26	121.1(8)
C12-N1-CN1	124.7(7)	O77-C17-N6	122.5(9)
C21-N1-CN1	115.6(7)	O77-C17-C27	121.6(10)
O22-C12-N1	120.9(7)	N6-C17-C27	115.8(9)
O22-C12-C22	120.7(7)	C17-C27-C37	114.0(9)
N1-C12-C22	118.3(7)	N7-C37-C27	113.4(9)
N2-C22-C12	111.4(6)	C18-N7-C37	129.6(10)
N2-C22-C3'2	108.7(7)	O88-C18-N7	117.2(11)
C12-C22-C3'2	108.5(6)	O88-C18-C28	122.3(12)
C32-C3'2-C22	114.2(7)	N7-C18-C28	120.3(10)
C3'2-C32-C42	112.6(9)	N8-C28-C38	114.9(8)
C3'2-C32-C52	110.4(8)	N8-C28-C18	107.8(8)
C42-C32-C52	110.1(9)	C38-C28-C18	113.1(9)
C13-N2-C22	124.6(7)	C28-C38-C58	113.7(9)
O33-C13-N2	122.4(8)	C28-C38-C48	106.7(11)
O33-C13-C23	120.2(8)	C58-C38-C48	108.8(9)
N2-C13-C23	117.4(8)	C68-C58-C38	114.3(11)
N3-C23-C33	111.6(7)	C19-N8-C28	121.2(8)
N3-C23-C13	111.4(7)	C19-N8-CN8	122.7(9)
C33-C23-C13	112.8(7)	C28-N8-CN8	115.9(9)
C14-N3-C23	120.9(7)	O99-C19-N8	120.4(8)
O44-C14-N3	123.3(8)	O99-C19-C29	117.4(9)
O44-C14-C24	117.7(8)	N8-C19-C29	122.3(9)
N3-C14-C24	118.7(8)	N9-C29-C19	109.8(8)
N4-C24-C34	111.8(8)	N9-C29-C39	107.4(11)
N4-C24-C14	112.8(7)	C19-C29-C39	109.5(8)
C34-C24-C14	110.9(7)	C59-C39-C29	120(2)
C15-N4-CN4	118.3(8)	C59-C39-C49	114.4(13)
C15-N4-C24	125.5(8)	C29-C39-C49	108.9(10)
CN4-N4-C24	115.0(8)	C110-N9-C29	123.7(8)
O55-C15-N4	121.5(9)	O1010-C110-N9	123.4(9)
O55-C15-C25	118.0(8)	O1010-C110-C210	120.0(9)
N4-C15-C25	120.6(8)	N9-C110-C210	116.6(9)
N5-C25-C35	114.3(8)	C110-C210-C310	115.0(8)
N5-C25-C15	105.8(8)	N10-C310-C210	112.2(9)
C35-C25-C15	112.1(7)	C111-N10-C310	119.8(9)
C45-C35-C25	108.3(9)	C111-N10-CN10	125.5(9)
C45-C35-C55	110.4(11)	C310-N10-CN10	114.7(8)
C25-C35-C55	108.4(9)	O1111-C111-N10	123.2(10)
C65-C55-C35	112.1(14)	O1111-C111-C211	120.0(9)
C16-N5-CN5	123.2(8)	N10-C111-C211	116.6(10)
C16-N5-C25	117.1(7)	N11-C211-C111	106.7(7)
CN5-N5-C25	119.6(8)	N11-C211-C311	113.6(7)

Bond angles

C111-C211-C311	109.5(8)
O11-C311-C411	107.0(8)
O11-C311-C211	108.2(7)
C411-C311-C211	113.3(8)
C112-N11-C211	125.2(8)
C11-O11-C311	117.0(8)
O1212-C112-N11	124.0(8)
O1212-C112-C212	120.0(9)
N11-C112-C212	116.0(9)
N12-C212-C3'12	113.4(7)
N12-C212-C112	108.6(7)
C3'12-C212-C112	111.5(7)
C312-C3'12-C212	114.2(8)
C3'12-C312-C512	111.4(10)
C3'12-C312-C412	111.1(10)
C512-C312-C412	109.6(10)
C113-N12-CN12	123.9(7)
C113-N12-C212	116.0(7)
CN12-N12-C212	118.3(7)
O1313-C113-N12	121.0(8)
O1313-C113-C213	118.8(8)
N12-C113-C213	120.0(8)
N13-C213-C313	113.5(8)
N13-C213-C113	106.3(7)
C313-C213-C113	113.0(7)
C213-C313-C413	112.3(8)
C213-C313-C513	109.2(8)
C413-C313-C513	109.7(9)
C114-N13-C213	123.9(8)
O1414-C114-N13	122.4(9)
O1414-C114-C214	121.4(9)
N13-C114-C214	116.1(9)
O1514-C214-C114	114.5(8)
C214-O1514-C014	112.5(8)

Symmetry transformations used togenerate equivalent atoms:
#1 x+1, y,z #2 x-1, y, z

Table A1.4. Anisotropic displacement parameters ($Å^2$ x 10^3) for **50**.
The anisotropic displacement factor exponent takes the form:
$-2\pi^2 [(ha^*)^2 U_{11} + ... + 2 h k a^* b^* U_{12}]$

	U11	U22	U33	U23	U13	U12
11	69(5)	37(4)	60(4)	-7(3)	-1(4)	0(4)
11	52(6)	27(6)	46(7)	-15(5)	0(6)	-10(5)
21	43(5)	37(5)	48(6)	-12(5)	-16(5)	-12(5)
3'	163(6)	42(6)	33(5)	-18(5)	-8(5)	-6(5)
31	79(8)	84(9)	57(7)	7(7)	-18(6)	-10(7)
41	105(11)	171(16)	103(11)	37(11)	-31(9)	23(11)
51	120(12)	147(14)	83(10)	64(10)	-3(9)	-29(11)
1	31(5)	40(4)	43(4)	-4(4)	-2(4)	-1(4)
N1	57(7)	83(8)	76(8)	-23(7)	-3(5)	-1(6)
22	42(4)	37(4)	47(4)	-1(3)	-5(3)	0(3)
12	21(6)	27(5)	36(5)	-4(4)	-4(4)	0(5)
22	24(4)	35(5)	36(5)	2(4)	-4(4)	0(4)
3'	246(5)	49(6)	42(5)	11(5)	-3(4)	10(5)
32	70(8)	34(6)	56(6)	5(5)	0(5)	21(5)
42	105(10)	63(8)	112(10)	11(7)	50(8)	25(7)
52	134(11)	28(6)	114(10)	17(6)	0(9)	12(7)
2	30(4)	32(4)	40(5)	-6(3)	-4(4)	-5(3)
33	40(4)	46(4)	36(3)	4(3)	8(3)	-16(3)
13	48(6)	15(4)	32(6)	-4(4)	0(5)	9(4)
23	35(5)	33(5)	26(5)	-5(4)	10(4)	-5(4)
33	46(6)	33(5)	58(6)	-3(5)	1(5)	-8(5)
3	32(4)	34(5)	34(4)	-7(3)	2(3)	-10(4)
O44	43(4)	36(4)	52(4)	-6(3)	3(3)	2(3)
C14	31(6)	37(6)	38(5)	-3(4)	6(4)	13(5)
C24	49(6)	34(6)	48(6)	1(4)	1(4)	0(5)
C34	60(7)	20(5)	164(12)	-6(6)	38(7)	-9(5)
N4	46(5)	37(4)	50(5)	-6(4)	-2(4)	-9(4)
CN4	64(7)	67(7)	47(6)	3(5)	-10(5)	-12(6)
O55	41(5)	51(4)	74(5)	3(3)	8(4)	-6(3)
C15	39(7)	26(5)	66(7)	-2(5)	11(6)	-7(4)
C25	53(6)	31(6)	50(6)	-3(4)	18(5)	-14(5)
C35	87(8)	45(7)	75(8)	6(6)	6(6)	-25(6)
C45	108(10)	54(7)	67(7)	13(6)	-20(7)	-34(7)
C55	312(25)	81(10)	42(8)	35(7)	70(11)	-12(13)
C65	184(18)	167(17)	106(12)	-32(12)	8(12)	-59(15)
N5	44(5)	42(5)	74(6)	-21(4)	29(4)	-17(4)
CN5	49(8)	88(9)	215(17)	-86(10)	71(9)	-44(7)
O66	47(4)	35(4)	49(4)	0(3)	15(3)	-17(3)
C16	49(7)	46(6)	35(5)	-8(4)	19(5)	-9(5)
C26	62(7)	34(6)	79(7)	-26(5)	34(6)	-27(5)
C36	141(12)	60(7)	62(8)	-31(6)	58(9)	-40(8)
C46	152(15)	103(11)	42(7)	2(7)	29(8)	-34(10)
C56	244(21)	55(8)	99(11)	-33(8)	131(13)	-40(11)
C66	222(27)	89(15)	680(72)	139(27)	315(40)	97(17)
N6	41(4)	28(4)	53(5)	3(4)	0(4)	-8(4)
O77	54(5)	59(5)	97(6)	-10(4)	-3(4)	-16(4)
C17	41(7)	53(8)	68(7)	-21(6)	-14(5)	-15(6)
C27	72(7)	42(6)	43(5)	-2(5)	10(5)	-3(5)
C37	75(7)	53(7)	59(7)	-14(6)	-4(6)	-7(6)

N7	57(6)	93(8)	115(9)	-67(7)	14(6)	-23(6)
O88	138(9)	89(11)	73(6)	-11(7)	9(6)-	99(9)
C18	55(8)	72(8)	65(8)	5(6)	0(7)-	23(7)
C28	59(6)	40(6)	56(7)-	13(5)	-4(5)	-18(5)
C38	102(9)	62(8)	62(7)	-4(6)	-15(7)	-33(7)
C48	236(20)	48(8)	72(9)	0(7)	-27(11)	-26(10)
C58	107(10)	60(8)	77(8)	2(6)	7(7)	-1(7)
C68	189(17)	115(12)	58(9)	-3(8)	8(10)	-58(12)
N8	42(5)	56(6)	49(5)	-3(4)	-8(4)	9(5)
CN8	69(8)	115(11)	96(10)	4(8)-	17(7)	29(8)
O99	54(5)	46(4)	71(5)	13(3)	20(4)	10(4)
C19	49(7)	37(6)	39(5)	-2(4)	-7(5)	15(6)
C29	90(8)	42(6)	37(6)	1(5)-	10(6)	16(6)
C39	319(27)	37(7)	44(8)-	10(6)	16(12)	4(11)
C49	325(27)	42(8)	86(10)	5(7)	-20(13)	20(12)
C59	422(36)	75(11)	62(10)	1(8)	29(15)	-86(17)
N9	51(5)	45(5)	57(6)	15(4)	-17(5)	-9(4)
O1010	41(5)	99(6)	59(5)	34(4)	-13(4)	-2(4)
C110	46(7)	48(6)	50(7)	24(5)-	11(5)	-11(5)
C210	52(6)	48(6)	90(8)	18(6)	-4(6)	-11(5)
C310	71(7)	57(7)	70(7)	14(6)	7(6)	-8(6)
N10	49(5)	47(5)	63(5)	23(4)	-5(4)	-3(4)
CN10	48(7)	67(7)	92(8)	21(6)	0(6)	-16(6)
O1111	44(5)	56(4)	108(6)	21(4)	-6(4)	-13(4)
C111	77(9)	36(7)	43(6)	-3(5)	11(6)	-8(6)
C211	62(6)	30(5)	54(6)	-2(5)	2(5)	-15(5)
C311	64(6)	30(5)	62(6)	3(5)	-2(5)	-13(5)
C411	126(10)	60(7)	55(7)	-8(6)	-16(7)	-24(7)
N11	44(5)	33(4)	43(5)	-5(4)	1(4)	-8(4)
O11	73(4)	33(4)	38(4)	-1(3)	5(4)	-5(3)
O1212	51(5)	94(6)	67(5)	-32(4)	3(4)	9(4)
C112	51(7)	40(6)	53(6)	-4(5)	8(6)	-12(5)
C212	50(5)	30(5)	38(6)	-12(4)	5(4)	-6(5)
C3'12	74(7)	29(5)	44(6)	2(5)	-3(5)	-3(5)
C312	90(9)	55(7)	59(7)	4(6)	3(6)	-3(7)
C412	116(11)	152(14)	65(8)	46(9)	3(8)	8(10)
C512	121(12)	120(12)	70(9)	9(8)	45(8)	53(10)
N12	43(5)	34(5)	37(4)	-1(4)	-13(3)	-2(4)
CN12	43(6)	50(6)	47(6)	12(5)	-9(5)	4(5)
O1313	55(5)	40(4)	60(4)	12(3)	-21(4)	-3(3)
C113	38(6)	38(6)	38(5)	1(5)	-15(5)	5(5)
C213	39(5)	29(5)	46(6)	-7(4)	-15(4)	4(4)
C313	58(6)	38(6)	66(7)	-5(5)	-17(5)	5(5)
C413	76(8)	44(7)	111(10)	16(6)	-14(7)	-18(6)
C513	86(8)	53(7)	83(8)	-9(6)	-2(7)	12(6)
N13	54(6)	46(5)	36(5)	1(4)	-11(4)	-9(4)
O1414	57(4)	53(4)	55(4)	3(3)	-12(4)	-3(3)
C114	51(7)	38(6)	50(7)	9(5)-	10(6)	-14(5)
C214	57(6)	54(6)	47(6)	0(5)	15(5)	-11(5)
O1514	85(5)	57(5)	64(4)	-1(4)	-23(4)	-2(4)
C014	124(11)	58(8)	82(9)	-6(7)	-4(8)	-1(7)

Table A1.5 Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^2 x 10^3) for **50**.

	x	y	z	U(eq)
H2A1	3738(7)	414(4)	1735(2)	51
H3'A1	1799(8)	-170(5)	1951(2)	55
H3'B1	2413(8)	412(5)	2132(2)	55
H3A1	3276(10)	-916(7)	2055(3)	88
H4A1	4517(49)	-621(23)	2450(27)	189
H4B1	4046(15)	120(41)	2462(25)	189
H4C1	4617(41)	-141(64)	2148(4)	189
H5A1	1777(50)	-1006(56)	2357(4)	175
H5B1	2179(78)	-437(14)	2595(18)	175
H5C1	2768(36)	-1144(46)	2579(19)	175
HNA1	4391(11)	-772(35)	1188(9)	108
HNB1	4616(19)	-847(30)	1559(10)	108
HNC1	4764(13)	-137(7)	1389(18)	108
H2B2	3243(6)	-1274(4)	1016(2)	38
H3'C2	2896(8)	-1829(5)	1535(2)	55
H3'D2	1720(8)	-1957(5)	1420(2)	55
H3B2	2492(9)	-2605(5)	977(2)	64
H4D2	4294(13)	-2237(20)	1073(21)	140
H4E2	4181(10)	-3029(30)	1028(18)	140
H4F2	4315(12)	-2719(48)	1375(4)	140
H5D2	1583(17)	-3142(24)	1420(19)	138
H5E2	2681(54)	-3221(28)	1598(9)	138
H5F2	2472(66)	-3617(7)	1275(11)	138
H0A2	1110(5)	-1522(3)	918(2)	41
H2C3	1255(6)	-1405(4)	120(2)	38
H3F3	-300(29)	-783(11)	126(10)	68
H3G3	654(10)	-304(9)	221(13)	68
H3H3	-120(35)	-583(18)	488(4)	68
H0B3	-350(6)	-1872(4)	477(2)	40
H2D4	219(7)	-3122(4)	812(2)	52
H3I4	38(38)	-4166(8)	508(18)	122
H3J4	1215(30)	-3895(21)	531(16)	122
H3K4	552(63)	-3783(15)	214(4)	122
HND4	-1898(41)	-2760(19)	136(5)	89
HNE4	-1792(45)	-3555(16)	172(4)	89
HNF4	-816(10)	-3123(33)	51(3)	89
H2E5	-691(8)	-2899(4)	1130(2)	54
H3L5	-2608(10)	-3565(5)	1334(3)	83
H4G5	-1571(52)	-4248(22)	977(7)	114
H4H5	-1565(52)	-4571(7)	1323(12)	114
H4I5	-572(11)	-4156(17)	1201(18)	114
H5G5	-630(19)	-3398(7)	1690(3)	174
H5H5	-1645(19)	-2954(7)	1763(3)	174
H6A5	-1488(102)	-3863(50)	2135(11)	229
H6B5	-1599(111)	-4386(12)	1850(21)	229
H6C5	-2566(21)	-3915(54)	1943(29)	229
HNG5	-3255(18)	-1716(7)	1285(30)	176
HNH5	-3259(17)	-2413(54)	1472(15)	176
HNI5	-3450(9)	-2401(55)	1098(15)	176

H2F6	-2377(9)	-971(5)	1283(2)	70
H3M6	-2054(13)	-1495(6)	1799(3)	105
H4J6	-588(15)	-1072(56)	2119(3)	148
H4K6	-112(30)	-737(32)	1807(20)	148
H4L6	-203(35)	-1534(24)	1833(22)	148
H5I6	-2335(18)	-30(7)	1747(4)	159
H6D6	-2780(333)	-76(173)	2225(71)	496
H6E6	-2806(347)	-875(147)	2203(82)	496
H6F6	-3665(21)	-440(310)	2020(12)	496
H0C6	-295(6)	-519(4)	1276(2)	49
H2G7	-530(8)	467(5)	729(2)	63
H2H7	78(8)	490(5)	1060(2)	63
H3N7	-628(9)	1581(5)	823(3)	75
H3O7	-1702(9)	1356(5)	986(3)	75
H0D7	36(8)	1815(6)	1297(3)	106
H2I8	221(8)	1872(5)	1865(2)	62
H3P8	-1817(11)	1831(6)	2146(3)	90
H4M8	-499(97)	815(22)	2076(3)	178
H4N8	-1155(44)	854(24)	2398(25)	178
H4O8	31(54)	1113(7)	2389(26)	178
H5J8	26(11)	2308(6)	2484(3)	98
H5K8	-887(11)	2803(6)	2373(3)	98
H6G8	-1038(76)	2522(48)	2917(9)	181
H6H8	-1202(86)	1754(21)	2822(16)	181
H6I8	-2061(22)	2301(64)	2725(9)	181
HNJ8	-1571(18)	3536(12)	1775(22)	140
HNK8	-2106(20)	2843(34)	1874(17)	140
HNL8	-1859(35)	2993(43)	1510(7)	140
H2J9	-351(9)	3967(5)	1532(2)	67
H3Q9	1445(20)	4113(6)	1943(3)	160
H4P9	1081(132)	5338(18)	1947(6)	227
H4Q9	286(46)	5210(30)	1661(34)	227
H4R9	1507(82)	5053(15)	1618(30)	227
H5L9	554(106)	3889(75)	2285(21)	279
H5M9	-499(41)	4131(107)	2115(4)	279
H5N9	283(150)	4670(35)	2266(24)	279
H0E9	1797(6)	3871(4)	1388(2)	61
H2K10	2419(8)	4319(5)	945(3)	76
H2L10	1618(8)	4660(5)	703(3)	76
H3R10	1300(9)	3496(5)	499(3)	79
H3S10	2365(9)	3855(5)	394(3)	79
HNM10	4057(10)	3095(36)	914(6)	103
HNN10	3919(8)	3670(11)	654(18)	103
HNO10	4058(10)	2907(28)	546(13)	103
H2M11	3471(8)	2014(4)	914(2)	58
H3T11	1485(8)	1372(4)	1076(2)	62
H4S11	1942(61)	570(18)	670(5)	120
H4T11	2983(12)	984(39)	592(8)	120
H4U11	1858(56)	1302(23)	518(5)	120
H0F11	2372(6)	2387(3)	1461(2)	48
H2N12	3114(7)	2885(4)	1856(2)	47
H3'E12	3577(8)	1800(5)	2066(2)	59
H3'F12	4779(8)	2025(5)	2073(2)	59
H3U12	4243(10)	2924(6)	2438(2)	82
H4V12	5190(12)	2033(45)	2665(18)	166
H4W12	4195(70)	1547(12)	2655(17)	166
H4X12	4223(73)	2161(36)	2898(3)	166

H5O12	2382(22)	2832(47)	2341(14)	156
H5P12	2648(12)	2802(49)	2710(13)	156
H5Q12	2380(21)	2129(9)	2522(25)	156
HNP12	6125(10)	3558(9)	1860(13)	70
HNQ12	5976(15)	2796(21)	1752(10)	70
HNR12	5779(9)	3004(29)	2113(5)	70
H2O13	5507(7)	4470(4)	1816(2)	46
H3V13	3873(8)	5139(5)	1472(2)	65
H4Y13	5339(52)	5755(25)	1343(5)	115
H4Z13	5825(28)	5687(22)	1690(16)	115
H4[A13	4836(27)	6163(7)	1631(18)	115
H5R13	3418(49)	5640(25)	1963(6)	111
H5S13	4422(10)	5326(40)	2134(4)	111
H5T13	3478(51)	4851(17)	2026(9)	111
H0G13	5243(7)	4264(4)	1167(2)	55
H2P14	7228(8)	4661(5)	886(2)	63
H2Q14	7857(8)	4034(5)	1024(2)	63
H0H14	6233(55)	2781(7)	748(15)	132
H0I14	6460(68)	2996(10)	1106(7)	132
H0J14	7406(17)	2933(13)	860(20)	132

Appendix Two

Publication

Journal of Natural Products (In print)

Theonellapeptolide IIIe, a New Cyclic Peptolide from the New Zealand Deep Water Sponge, *Lamellomorpha strongylata.*

Shangxiao Li,[+] Eric J. Dumdei,[+] John W. Blunt,*[+]

Murray H. G. Munro,*[+] Ward T. Robinson,[+] and Lewis K. Pannell.[‡]

Department of Chemistry, University of Canterbury, Christchurch, NEW ZEALAND, AND Laboratory of Bioorganic Chemistry, NIDDK, NIH, Bethesda, MD 20892-0805, USA.

The structure, stereochemistry and conformation of theonellapeptolide IIIe (**1**), a new 36-membered ring cyclic peptolide from the New Zealand deep-water sponge *Lamellomorpha strongylata*, is described. The sequence of the cytotoxic peptolide was determined through a combination of NMR and MS/MS techniques and confirmed by X-ray crystal structure analysis, which, with chiral HPLC, established the absolute stereochemistry.

Since 1982, more than 5,000 specimens of marine organisms have been collected from the oceans around New Zealand in a continuous endeavour to find bioactive natural products. Over that period, compounds with a wide degree of structural diversity and bioactivity have been discovered.[1-3] To further explore the potential pharmaceutical resources of deep-water marine organisms, an expedition to the Chatham Rise, 200 km off the East Coast of the South Island, was carried out in early 1995. Preliminary assays on the crude extracts indicated that deep-water organisms are a rich source of bioactive compounds.[4]

A bioassay-guided separation of the extract from *Lamellomorpha strongylata* from the collection resulted in the isolation of three distinct classes of cytotoxic compounds. We have previously reported the characterisation of calyculins A, B, E, and F, calyculinamides A and B, and swinholide H,[4] which comprise two of the classes of cytotoxins from this sponge. We now present the structure of theonellapeptolide IIIe (**1**), the major component in the third cytotoxic fraction from the extract.

Results and Discussion

A frozen sample of *L. strongylata* (1 kg) was homogenised, extracted with MeOH:DCM (1:1) and subjected to rounds of C18, DIOL and LH-20 chromatography, as described previously.[4] LH-20 chromatography (MeOH:DCM; 1:1) of the more polar DIOL fractions resulted in a fraction with weak P388 activity (IC_{50} 25 μg/mL) which eluted just prior to the calyculinamides A and B. The peptidic nature of the fraction was indicated by both the ^1H NMR (doublets between δ 7.39-9.00) and the ^{13}C NMR (resonances between δ 169-176) spectra. Purification, by repeated RPHPLC, yielded five new active compounds, theonellapeptolides IIIa, IIIb, IIIc, IIId and IIIe, in order of elution.

Theonellapeptolide IIIe (**1**), the most abundant peptide, exhibited an MH^+ ion at *m/z* 1419 in the LRFAB mass spectrum. The molecular formula $C_{71}H_{127}N_{13}O_{16}$ was determined by HRFABMS (*m/z* 1550.8577, $[M+Cs]^+$, Δ -0.1). A negative result against ninhydrin suggested that the N-terminus was blocked, or part of a cyclic peptide, while the observation of a lactone carbonyl stretch (1734 cm^{-1}) in the IR spectrum, coupled with the unsaturation requirements of the molecular formula, indicated that **1** was most likely a peptolide rather than a peptide. The ^1H NMR spectrum showed seven amide proton signals (δ 9.00, 8.76, 8.43, 8.19,

7.65, 7.45, 7.39) and six *N*-methyl singlets (δ 3.33, 3.30, 3.24, 3.19, 2.75, 2.69). Searching of the MarinLit[5] database revealed structural similarity of this new compound to the known theonellapeptolides.[6-8]

Amino acid analysis by GC-MS of the *N*-trifluoroacetamide-*n*-butyl ester derivatives[9,10] formed after complete acid hydrolysis of the peptolide confirmed thirteen amino acids: Ala, *N*-MeAla, β-Ala, *N*-Meβ-Ala, Ile, *N*-MeIle (x2), Leu, *N*-MeLeu (x2), Thr and Val (x2). An additional carbonyl resonance in the ^{13}C NMR spectrum, as well as signals corresponding to isolated methylene (δ 3.86, s, 2H) and methoxy (δ 3.16, s, 3H) functionalities in the ^{1}H NMR spectrum, indicated the probable existence of a 2-methoxyacetyl group, blocking the *N*-terminus of the peptolide.

To determine the amino acid sequence, the ring-opened methyl ester (**2**) was prepared by methanolysis. Detailed examination of MS/MS data from the FAB spectrum of **2** established the sequence shown in Figure 1. Of particular note was the observation of daughter ions at *m/z* 299, 172 and 100 in the link scan from the ion at *m/z* 400. These ions corresponded to the loss of Thr, *N*-MeLeu or *N*-MeIle and 2-methoxyacetyl (from *m/z* 172) respectively, thus confirming the *N*-terminus region.

In analysing FABMS data it is not possible to distinguish between isobaric amino acids, such as Ala and β-Ala, *N*-MeAla and *N*-Meβ-Ala, Leu and the Ile's, and *N*-MeLeu and the *N*-MeIle's Therefore, the suggested sequence had to be confirmed by other methods. This confirmation of the amino acid sequence in **1** was accomplished by ^{1}H-^{1}H decoupling experiments and extensive analyses of the TOCSY, COSY, ROESY, HSMQC[11] and HMBC NMR spectra. The spin systems (see

Table 1) for each of the individual amino acid partial structures, except for *N*-MeAla (residue 10) and *N*-MeLeu (residue 13), were defined by COSY and 2D TOCSY experiments. The *N*-MeAla subunit was subsequently defined by specific ^{1}H-^{1}H decoupling experiments, which showed that the unassigned methyl doublet at δ 1.44 was coupled into an α-methine multiplet at δ 5.20. For the *N*-MeLeu (residue 13) examination of the 2D TOCSY spectrum indicated that an unassigned methine signal at δ 3.74 gave correlations to ^{1}H resonances at δ 2.05, 1.06 and 0.98, as well as a weak correlation to δ 1.66, but the δ 3.74 methine resonance was **not** correlated to any carbons in the HMBC spectrum. However, a correlation between the protons at δ 3.74 and 2.05 in the COSY spectrum, and correlations to δ 2.05 from both δ 3.74 and 1.66 in 1D TOCSY experiments were observed, which suggested the signal at δ 3.74 was indeed the α-proton for an *N*-MeLeu. A ^{1}H *N*-methyl signal (δ 3.30) and the ^{1}H resonance at δ 2.05 which were correlated with a carbon (δ 63.0) resonance in the HMBC spectrum supported the given assignments.

Correlations between NH, *N*-Me or α-protons to carbonyl carbons in the HMBC spectrum afforded much of the connectivity data. For example, the correlations from H3 of Thr (residue 3) to the carbonyl resonance of *N*-MeLeu (residue 13) suggested that the Thr and *N*-MeLeu were linked by an ester bond, and so established the ring system of the peptolide. Complete assignments for NMR resonances, based on 1D and 2D NMR experiments, are summarised in Table 1. Correlations observed in the ROESY experiment confirmed the sequence of peptide bonds deduced in the HMBC experiment. Interestingly, the ROESY spectrum also showed cross peaks between the NH of β-Ala (residue 7) and the *N*-Me of *N*-MeIle (residue 6), between the NH's of Ala and Leu (residues 11

and 12), and between the NH of Ala and the *N*-Me of *N*-MeAla (residues 11 and 10), indicating *syn* orientations about *N*-MeIle/β-Ala, Ala/Leu and Ala/*N*-MeAla peptide bonds. The same *syn* orientations were also observed in the X-ray analysis (*vide infra*) suggesting at least comparable geometries for the solution and solid phases for **1**. The peptide sequence deduced from analysis of the NMR data was consistent with the MS/MS analysis.

At this point, the stereochemistry of the component amino acids had not been addressed. This was resolved by a combination of X-ray crystallography and chiral HPLC. The structure was solved by direct methods and refined using the SHELXTL system of programs and all the unique intensity data.[12,13] The best converged conventional R-factor was 12.9% for 12,029 reflections with $F_o > 4\sigma(F_o)$. All numerical data came from programs in the SHELXTL system.[12,13]

In addition to the theonellapeptolide molecule shown (Figure 2), there were at least 21 sites for the oxygen atoms of water molecules, 14 of which were at distances appropriate for H-bonding to carbonyl O and N atoms of the main chain. These water molecules are labelled in the Supplementary Tables. The other 7 water molecules participated in water-to-water H-bonding. The atomic displacement parameters were consistent with some disorder amongst these water molecules. Six sites, involved in inter-atomic distance interactions between 1.6 and 2.0 Å, must be only partially occupied. These features have so far hindered attainment of better conventional agreement factors. Nevertheless, the structure is fully characterised by this study.

Because the X-ray analysis gave only the relative stereochemistry of the cyclic peptolide the absolute configuration of at least one of the

individual amino acid residues needed to be determined. This was achieved by chiral HPLC analysis of the constituent amino acids.[14] A sample of the acid hydrolysis product was derivatized with dansyl chloride (Dan-Cl).[15] The dansylated derivatives of amino acids were coinjected with standard Dan-L-Thr (2*S*,3*R* stereogenic centres) and Dan-D,L-Thr respectively. The Dan-Thr from the hydrolysate was found to co-elute with standard Dan-L-Thr, establishing 2*S*,3*R* stereochemistry at the Thr stereogenic centres. From the relative stereochemistries of theonellapeptolide IIIe (**1**), established by X-ray crystallography, it was then possible to assign the stereochemistry at all stereogenic centres. Thus, the chiral amino acids of **1** could be assigned as: L-Val, L-Thr, D-Val, D-*allo*-Ile, L-*N*-MeIle (x 2), L-*N*-MeAla, L-Ala, D-Leu and D-*N*-MeLeu (x 2).

Based on the general A - B bond length (2.79 ± 0.12 Å) for N-H...O bonds, three intramolecular H-bonds can be assumed; (crystallographic numbering [residues,distance]) N9 - O13 [Val (residue 5)-Val (residue 1) 2.89 Å], N11 - O9 [Thr (residue 3)-Val (residue 5) 2.85 Å], and N2 - O6 [Leu (residue 8)-*allo*-Ile (residue 12) 2.91 Å], with a further, weaker H-bond that should also be taken into consideration, N6 - O2 [*allo*-Ile (residue 8)-Leu (residue 12), 3.08 Å]. The latter three of these H-bonds are across the peptolide ring and stabilise the complex conformation found in **1**. The presence of a β-Ala and a *N*-Meβ-Ala allow for considerable flexibility in the conformation of the cyclic peptolide ring, which despite this flexibility, can be considered as an antiparallel array of two strands in which the turns are secured at one end by a β-Ala (residue 4) and at the other by a *cis*-peptide bond [*N*-MeIle (residue 9)/*N*-MeAla (residue 10)]. The resulting conformation is a triple loop,[12] which is more clearly depicted in the backbone projection

diagram (Figure 3). The two major loops are defined by the two H-bonds between the *allo*-Ile and the Leu (residues 8 and 12) which forms an antiparallel β-pleated sheet structure. The third loop is defined by weak NH...O associations from the NH of the β-Ala (residue 7) to Thr (residue 3) and MeLeu (residue 13) {N7 - O1 [β-Ala (residue 7)-(Thr (residue 3), 3.28Å], N7 - O11 [β-Ala (residue 7)-(*N*-MeLeu (residue 13), 3.31Å]}. That portion of the peptide outside the macrocycle is static, being held in place by two H-bonds from Val (residue 5) to Thr (residue 3) and Val (residue 1). The overall conformation of **1** resembles a "taco shell" (Figure 2) with all the larger aliphatic groups aligned along the lip of the "taco shell", while the base is composed of the smaller amino acids (Ala, β-Ala, *N*-Meβ-Ala).[a]

The reported theonellapeptolides I and II differ mainly in variation among the aliphatic side chains. For example, theonellapeptolide Ie,[7] which has the same molecular formula as **1**, has a similar peptolide backbone, but varies in the fifth, ninth, eleventh and thirteenth amino acid residues from N-terminus, corresponding to substitutions of D-Val for D-Leu, L-*N*-MeIle for L-*N*-MeVal, L-Ala for β-Ala, and D-*N*-MeLeu for D-*N*-Me*allo*-Ile respectively. These modifications result in a 36-, rather than 37-membered peptolide ring for **1**. An X-ray structure has been reported in a conference abstract[16] for theonellapeptolide Id. Theonellapeptolide Id differs from Ie only in substitution of a β-Ala for a *N*-Meβ-Ala. Despite the substitution of five amino acid substitutions and the difference in ring size there is a quite remarkable similarity in the overall conformation and H-bonding patterns between theonellapeptolides IIIe (**1**) and Id, which suggests that the theonellapeptolide family quite probably have similar tertiary structures.

---

[a] Best seen in stereo projection. See Supplementary Figure.

Theonellapeptolide IIIe exhibited modest cytotoxicity (7.4 μg/mL) against the P388 cell line. This sample was also tested for antiviral and anti-HIV activities. However, only cytotoxicity and no protection was observed. Other reported biological activties for the theonellapeptolide family are ion-transport properties for theonellapeptolide Id (Na^+, K^+, and Ca^{2+}) and Ie (Na^+, K^+ across human erythrocyte membranes).[6-8] It is pertinent to note that the theonellapeptolide III compounds were isolated from a *Lamellomorpha* sp. sponge, which is in a different order from *Theonella swinhoei,* from which both the theonellapeptolides I and II were isolated.[6-8]

As it is rare to find so many interesting structures and associated bioactivities in one species, further chemical and biological studies are currently under way on *Lamellomorpha strongylata.* In particular, the question of what symbiotic microorganisms are present, and their role in the production of the bioactive metabolites will be addressed.

Experimental Section

General Experimental Procedures. MPs were determined on an Electrothermal melting point apparatus and are uncorrected. Optical rotations were measured on a JASCO J-20 recording spectropolarimeter. IR spectra were recorded on a Shimadzu FTIR-8201PC infrared spectrophotometer. GC-MS spectra were measured on a Kratos MS80RFA mass spectrometer equipped with a Carlo Erba MFC500 GC. FAB and high resolution FAB mass spectra were recorded on a JEOL SX102A spectrometer. ^1H and all 2-D NMR spectra were recorded on a Varian UNITY 300 NMR spectrometer. ^{13}C NMR spectra were recorded on a Varian XL 300 NMR spectrometer.

Extraction and Isolation. The formal description of *Lamellomorpha strongylata* and the extraction and separation of the theonellapeptolide fraction from the calyculin and swinholide fractions has been reported previously.[4] Final purification of the peptolides was carried out by reverse-phase HPLC on a Philips PU4100 using an ODS column (4.6 mm x 25 cm) eluting with H_2O (0.1% TFA)/MeOH (0.1% TFA) (11:89). A limited sample (2 mg) of the crude peptide was injected in each run. Cutting the appropriate peak in multiple runs gave pure theonellapeptolide IIIe (**1**) (94 mg).

Theonellapeptolide IIIe (1): Colorless plates (acetone/H_2O: 1:1); mp 184-186 ^{o}C; $[\alpha]^{22}_{D}$ -48.6^{0} (c 1.0, MeOH); IR ($CHCl_3$) ν_{max} 3352, 3006, 2966, 2935, 1734, 1672, 1628, 1533, 1468, 1173 cm^{-1}; ^{1}H NMR (CD_3OD/C_6D_6, 2:8) see Table 1; ^{13}C NMR (CD_3OD/C_6D_6, 2:8) see Table 1; HRFABMS, *m/z* 1550.8577 $[M+Cs]^+$, (calcd. for $C_{71}H_{127}N_{13}O_{16}Cs$ 1550.8578).

Acid Hydrolysis of 1. Theonellapeptolide IIIe (**1**) (1 mg) was placed in a hydrolysis tube and HCl (6 M) (1 mL) was added. After sealing under N_2, the tube was heated at 110 ^{o}C (17 h). The acid solution was evaporated to dryness under vacuum to give the product (1 mg).

Methanolysis of 1. Theonellapeptolide IIIe (**1**) (9 mg) was dissolved in dry MeOH (2 mL) with NaOMe/MeOH (4%) (0.6 mL). The MeOH solution was stirred at room temperature (3 h), before pouring into ice water and extracted with EtOAc (3 x 1 mL). Evaporation of the solvent gave a product which was purified by reverse-phase HPLC [H_2O (0.1% TFA)/MeOH; 18:82] to give the methanolysis product **2** (6 mg). HRFABMS *m/z* 1582.8860 $[M+Cs]^+$, (calcd. for $C_{72}H_{131}N_{13}O_{17}Cs$ 1582.8840).

Amino Acid Analysis by GC-MS The *N*-trifluoroacetamide-*n*-butyl esters of the amino acids were prepared under standard conditions.[10] The acid hydrolysate of **1** (1 mg) was transferred to a Reacti-vial (1 mL) and dried under vacuum before HCl-*n*-BuOH (2 M) (400 μL) was added. The vial was sealed with a Teflon-lined cap and after sonication (15 s), was heated at 110 °C in an oil bath (10 min), followed by further sonication (10 s) and heating (20 min). After cooling the solvent was evaporated under vacuum. A trifluoroacetic anhydride-DCM mixture (750 μL) (1:2) was added to the sample, which was then heated (150 °C for 5 min). The remaining reagent and solvent were removed under N_2 at 50 °C and the product then dissolved in dry DCM (600 μL) for GC analysis. The GC-MS analysis was performed on a DB Wax column (30 m x 0.25 mm) with an initial column temperature of 50 °C which was held for 15 min post-injection and then increased to 130 °C at 1 °/min, to 250 °C at 23 °/min and finally held at 250 °C for 20 min. The individual amino acids were identified on the basis of retention time against standards, and by analysis of the MS data.

X-ray Structure Determination of Theonellapeptolide IIIe (1). A colourless crystal of theonellapeptolide IIIe (**1**), obtained as a polyhydrate from acetone-water (1:1), was used in the data collection. Data reported here were derived from an irregular chunk of dimensions (0.6 x 0.4 x 0.15) mm^3. Crystal data: formula $C_{71}H_{127}N_{13}O_{16}/(H_2O)_{18}$, FW=1418/(324); orthorhombic, $P2_12_12_1$, a=12.507(3), b=19.573(4), c=41.416(8) Å, V=10,139(8) $Å^3$, Z=4, ρ (calculated) = 1.142 Mg/m^3, T = -70 °C.

The data collection nominally covered a hemisphere. Intensities of 18,590 unique reflections were collected on a Siemens SMART area

detector system, fitted with a nitrogen low-temperature gas flow device, using MoKα (λ=0.71073 Å) X-radiation. These data were processed using program SAINT which corrects for Lp effects and crystal decay for the duration of the experiment. All crystallographic data is reported in the Supplementary Tables.

Determination of Absolute Configuration. A sample of the acid hydrolysate of theonellapeptolide IIIe **1** (1 mg) was dissolved in aq. Li_2CO_3 (0.04 M) (3 mL) and reacted with dansyl chloride (Dan-Cl 2.2 mg/AcCN 1mL) (2.3 mL)[15] at room temperature (2 h). The solvent was removed under vacuum. The Dan-amino acids were further purified on a DIOL cartridge with DCM-ether-HOAc (5:1:1) as eluent. The first fraction (0.8 mL) was collected and dried under N_2 (0.8 mg). The Dan-derivatives were analysed on a reverse-phase HPLC column (ODS, 4.6 x 220 mm) run isocratically using L-Val (4 mM)-copper acetate (2 mM)-sodium acetate (0.3 M) (pH 7) as chiral eluent (1 mL/min). Standard Dan-L-Thr and Dan-DL-Thr were coinjected with the dansyl derivatives.

Supplementary Material Available: Copies of tables of final fractional atomic coordinates and the full list of bond lengths and angles from the X-ray crystallographic study of **1**, stereoprojection figure. This material is contained in libraries on microfiche, immediately follows this article in the microfilm version of the Journal, and can be ordered from the ACS. See any current masthead page for ordering information.

Acknowledgments. We thank the University of Canterbury, PharmaMar SA, Canterbury Medical Research Foundation, and the New Zealand Lotteries Board (Scientific) for financial support; Dr M. R. Boyd

and Dr M. J. Currens (NCI/DTP/LDDRD) for HIV assays, Ms G. Ellis (University of Canterbury) for other biological assays, and Mr. B. M. Clark (University of Canterbury) for mass spectral analyses.

References

(1) Perry, N. B.; Blunt, J. W.; Munro, M. H. G. *Tetrahedron* **1988**, *44*, 1727-1734.

(2) Perry, N. B.; Blunt, J. W.; Munro, M. H. G.; Thompson, A. M. *J. Org. Chem.* **1990**, *55*, 223-227.

(3) Litaudon, M.; Hart, J. B.; Blunt, J. W.; Lake, R. J.; Munro, M. H. G. *Tetrahedron Lett.* **1994**, *35*, 9435-9438.

(4) Dumdei, E. J.; Blunt, J. W.; Munro, M. H. G.; Pannell, L. K. *J. Org. Chem.* **1997**, *62*, 2636-2639.

(5) Marine Chemistry Group, University of Canterbury, New Zealand; *MarinLit, A Database of the Literature on Marine Natural Products.*

(6) Kobayashi, M.; Lee, N. K.; Shibuya, H.; Momose, T.; Kitagawa, I. *Chem. Pharm. Bull.* **1991**, *39*, 1177-1184.

(7) Kitagawa, I.; Lee, N. K.; Kobayashi, M.; Shibuya, H. *Chem. Pharm. Bull.* **1987**, *35*, 2129-2132.

(8) Kitagawa, I.; Ohashi, K.; Kawanishi, H.; Shibuya, H.; Shinkai, K.; Akedo, H. *Chem. Pharm. Bull.* **1989**, *37*, 1679-1681.

(9) Leimer, K. R.; Rice, R. H.; Gehrke, C. W. *J. Chromatogr.* **1977**, *141*, 121-144.

(10) Zumwalt, R. W.; Desgres, J.; Kuo, K. C.; Pautz, J. E.; Gehrke, C. W. *J. Assoc. Off. Anal. Chem.* **1987**, *70*, 253-262.

(11) Torres, A. M.; Nakashima, T. T.; McLung, R. E. D. *J. Mag. Res. Ser.* A **1993**, *102*, 219-227.

(12) Sheldrick, G. M. *J. Appl. Crystallogr.* In Press.

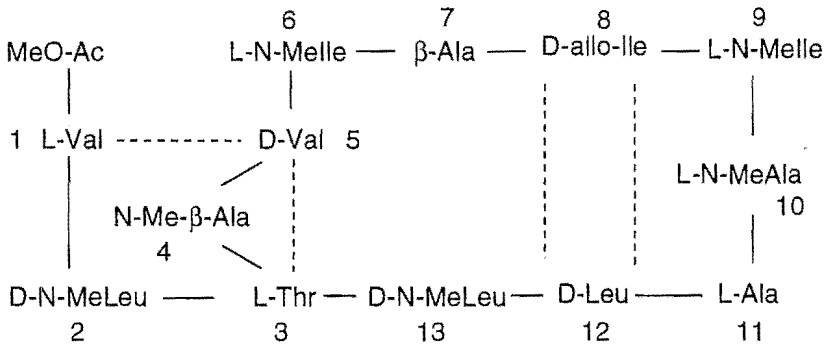
(13) Sheldrick, G. M. *Acta. Crystallogr., Sect. A.* **1990,** *46,* 467.

(14) Armani, E.; Barazzoni, L.; Dossena, A.; Marchelli, R. *J. Chromatogr.* **1988**, *441*, 287-298.

(15) Weinstein, S.; Weiner, S. *J. Chromatogr.* **1984**, *303*, 244-250.

(16) Bernardinelli, G.; Jefford, C. W.; Sakai, R.; Higa, T. *Proceedings, Seventh International Symposium on Marine Natural Products*, July, 1992, Capri, Italy. P29.

Scheme 1



------- H-bond. Residue numbers shown for each amino acid.

Theonellapeptolide IIIe (**1**)

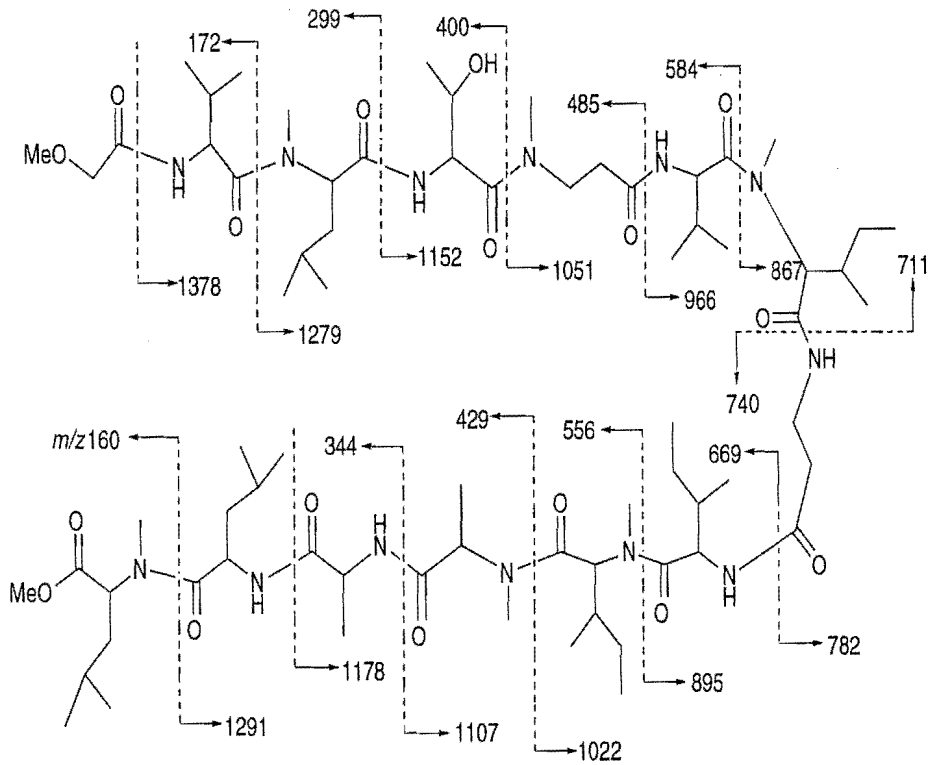


Figure 1 FAB MS/MS Fragmentation of **2** (protonated)

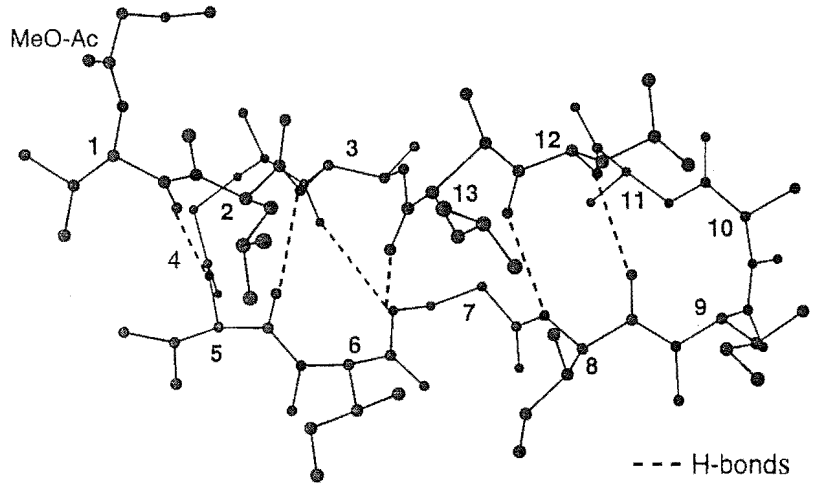


Figure 2. Perspective View of Theonellapeptolide IIIe (**1**)

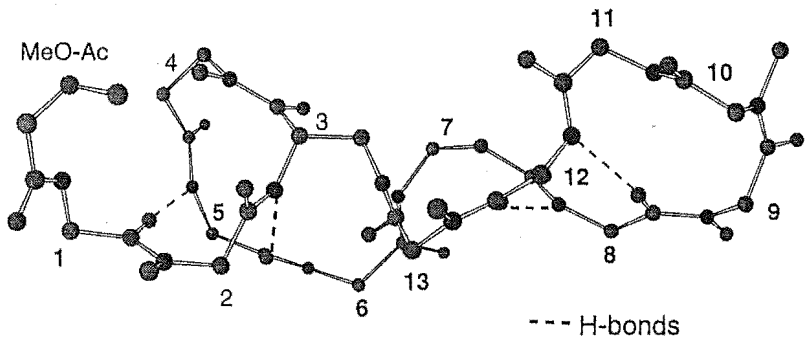


Figure 3. Backbone Projection of Theonellapeptolide IIIe (**1**)

Table 1. Correlated ^{1}H and ^{13}C Spectral Data for Theonellapeptolide IIIe (**1**)

Amino acid (residue #)		δ^a ^{13}C(# H)^b	^{1}H (mult,J(Hz))^c	HMBC^d	ROESY^e
L-Val (1)	1	174.4 (0)			
	2	54.5 (1)	5.03 (m)	174.4, 170.4, 32.1, 19.6, 17.6	3.19, 2.01, 0.98
	3	32.1 (1)	2.01 (m)	54.5, 19.6, 17.6	5.03, 3.19, 0.98, 0.91
	4	17.6 (3)	0.91	19.6	5.03, 2.01
	5	19.6 (3)	0.98	17.6	5.03, 2.01,
	NH		7.45 (d, 9.3)	170.4	5.03, 3.86, 2.72, 0.98
D-*N*-MeLeu (2)	1	172.8 (0)			
	2	56.4 (1)	5.28 (m)	174.4, 172.8, 38.0, 31.9, 25.9	9.0, 8.19, 2.05, 1.48
	3	38.0 (2)	2.05 (m)	56.4, 25.9	1.48, 0.94
			1.48 (m)	25.9, 56.4	5.28
	4	25.9 (1)	1.48 (m)	24.1	3.19, 2.05, 0.94
	5	10.7 (3)	0.94	38.0, 25.9, 24.1	5.28, 3.19, 2.05, 1.48
	6	21.4 (3)	1.06	38.0, 25.9	2.05
	NCH_3	31.9 (3)	3.19 (s)	174.4, 56.4	5.03, 2.01, 1.48, 0.98, 0.94
L-Thr (3)	1	169.2 (0)			
	2	53.7 (1)	5.00 (m)	172.8, 169.2, 18.2	7.65, 2.75, 1.31
	3	71.2 (1)	5.47 (dd, 8.7, 6.3)	172.3, 53.7	7.65, 3.53, 2.75, 1.31
	4	18.2 (3)	1.31 (d, 6.3)	71.2, 53.7	5.47, 5.00
	NH		9.00 (d, 6)		5.47, 5.28, 5.00
N-MeβAla (4)	1	172.9 (0)			
	2	34.7 (2)	2.30 (m)	172.9, 45.4	
			2.51 (m)	172.9, 45.4	8.19, 4.58, 2.30
	3	45.4 (2)	2.72 (m)		
			4.58 (m)	34.7	2.72, 2.51, 2.30
	NCH_3	35.0 (3)	2.75 (s)	169.2, 45.4	7.65, 5.47, 5.00, 2.51
D-Val (5)	1	174.0 (0)			
	2	55.4 (1)	5.06 (m)	172.9, 174.0, 31.0, 18.3	8.19, 3.24, 2.06, 0.99, 0.98
	3	31.0 (1)	2.06 (m)	55.4, 18.3	8.19, 5.06, 3.24, 0.98
	4	18.3 (3)	0.98	31.0	8.19, 5.06, 2.06
	5	19.7 (3)	0.99	31.0	
	NH		8.19 (d, 8.7)	172.9, 55.4	5.28, 5.06, 2.51, 2.06, 0.98
L-*N*-MeIle (6)	1	170.8 (0)			
	2	61.4 (1)	5.22 (m)	174.0, 170.8, 33.1, 31.6, 16.1	7.65, 3.24, 2.29, 1.13
	3	33.1 (1)	2.29 (m)		3.24, 1.43, 1.13
	4	16.1 (3)	1.13 (d, 6.3)	61.4, 33.1, 25.3	5.22, 2.29, 1.43
	5	25.3 (2)	1.43 (m)		2.29, 0.96
	6	21.8 (3)	0.96	25.3	
	NCH_3	31.6 (3)	3.24 (s)	174.0, 61.4	5.06, 2.29, 2.06
β-Ala (7)	1	172.5 (0)			
	2	35.9 (2)	2.51 (m)	172.5	8.76, 5.47, 3.53, 1.31
			2.69 (m)		8.76, 3.53
	3	35.8 (2)	3.53 (m)	172.5, 170.8	7.65, 5.47, 4.54, 2.51
			4.54 (m)	172.5, 170.8	3.53, 2.69
	NH		7.65 (d, 5.4)	170.8	5.47, 5.22, 5.00, 4.54, 3.53, 3.24
D-*allo*-Ile (8)	1	175.9 (0)			
	2	53.6 (1)	5.38 (m)	175.9, 172.5, 38.8, 15.5	8.76, 3.33, 1.75, 1.55, 0.98
	3	38.8 (1)	1.75 (m)	53.6, 26.9, 15.5, 12.4	5.38, 3.33, 1.04, 0.98
	4	15.5 (3)	0.98		
	5	26.9 (2)	1.22 (m)	53.6, 38.8, 15.5, 12.4	1.55
			1.55 (m)	53.6, 38.8, 12.4	5.58, 1.22, 1.04
	6	12.4 (3)	1.04		
	NH		8.76 (d, 9.3)	172.5	5.38, 2.69, 2.51, 0.98
L-*N*-MeIle (9)	1	172.7 (0)			
	2	55.5 (1)	5.18 (m)	175.9, 172.7, 33.8, 31.7, 29.7	3.33, 2.69
	3	33.8 (1)	2.28 (m)		3.33, 0.78
	4	15.8 (3)	0.78 (d, 12)	55.5, 33.8, 25.6	2.69, 2.28, 1.42,
	5	25.6 (2)	1.42 (m)	15.8	0.82
	6	9.8 (3)	0.82 (dd, 9.6, 6.3)	33.8, 25.6	2.28, 1.42
	NCH_3	31.7 (3)	3.33 (s)	175.9, 55.5	5.38, 5.18, 2.28, 1.75, 1.42
L-*N*-MeAla (10)	1	171.1 (0)			
	2	57.5 (1)	5.20 (m)	172.7, 171.7, 29.7, 15.0	2.69, 1.44
	3	15.0 (3)	1.44 (d, 6.9)	171.1, 57.5	5.20, 2.69
	NCH_3	29.7 (3)	2.69 (s)	172.7, 57.5	5.20, 5.18, 1.44, 0.78
L-Ala (11)	1	174.7 (0)			
	2	52.3 (1)	4.51 (m)	174.7, 17.9	1.29
	3	17.9 (3)	1.29 (d, 7.5)	174.7, 52.3	7.39, 4.51
	NH		7.39 (d, 6.3)	171.1, 52.3, 17.9	8.43, 4.51, 2.69, 1.29
D-Leu (12)	1	174.9 (0)			
	2	48.8 (1)	5.15 (m)	174.9, 40.8	8.43, 3.30, 1.87
	3	40.8 (2)	1.34 (m)		1.87, 0.98
			1.87 (m)	174.9, 48.8, 25.4	8.43, 5.15, 1.34
	4	25.4 (1)	1.87 (m)	21.3	1.04, 0.98
	5	21.3 (3)	1.04	40.8, 25.4, 23.9	1.87
	6	23.9 (3)	0.98	40.8, 25.4, 21.3	
	NH		8.43 (d, 8.4)	174.7, 48.8	7.39, 5.15, 1.87, 1.34
D-*N*-MeLeu (13)	1	172.3 (0)			
	2	63.0(1)	3.74 (m)		3.30, 2.05, 1.06
	3	38.3 (2)	2.05 (m)	172.3, 63.0, 25.6, 23.8, 22.5	3.74, 1.06, 0.98
	4	25.6 (1)	1.66 (m)	38.3, 23.8	3.30
	5	23.8 (3)	0.98	38.3	
	6	22.5 (3)	1.06	38.3, 25.6	3.74, 2.05, 1.66
	NCH_3	38.8 (3)	3.30 (s)	174.9, 63.0	5.15, 3.74, 1.66
$MeOCH_2CO$	1	170.4 (0)			
	2	72.1 (2)	3.86 (s)	170.4, 59.0	7.45, 3.16
	OCH_3	59.0 (4)	3.16 (s)	72.1	3.86

[a] Referenced to residual solvent C_6D_6 δ_H=7.27, δ_C=128.4. [b] ^{13}C spectra recorded on a Varian XL300 at 75 MHz. Number of attached H determined by DEPT. [c] ^{1}H spectra recorded on a Varian UNITY 300 at 300 MHz. [d] J_{nxh} = 8.3, 5.0 arrayed mixing times. [e] A mixing time of 0.15s was used.